

21世纪全国高等院校**财经管理**系列实用规划教材



统计学

（第2版）

主　编／阮红伟



- 关注前沿并贯穿最新行业规范
- 提供大量实际经济生活的案例
- 减少高深理论推导和公式论证
- 增加“习题与实训”的比例







21 世纪全国高等院校财经管理系列实用规划教材

统计学（第 2 版）

主　编　阮红伟
主　审　张[illegible]景

北京大学出版社
PEKING UNIVERSITY PRESS

内容简介

本书是一本统计学入门读物，主要内容包括统计概论、统计调查、统计整理、综合指标、时间数列、统计指数、抽样推断、相关分析与回归分析等。

本书结构模块包括理论目标、能力目标、做一做、想一想、正文、本章小结、应用与拓展、习题与实训。习题与实训的参考答案为电子资料，读者可从 www.pup6.com 自行下载参考。

本书内容难度适中，实用性、操作性强，案例丰富，通俗易懂，让学生在获得先进、系统的学术知识的同时，会有“统计就在你我身边”的亲切感。

本书适合作为普通高等院校（应用型本科、高职高专）、成人高校、民办高校及本科院校举办的二级职业技术学院经济类、管理类专业教材，也可供社会培训机构使用或自学者参考。

图书在版编目(CIP)数据

统计学/阮红伟主编．—2版．—北京：北京大学出版社，2014.3
（21世纪全国高等院校财经管理系列实用规划教材）
ISBN 978-7-301-23854-7

Ⅰ.①财…　Ⅱ.①阮…　Ⅲ.①统计学—高等学校—教材　Ⅳ.①C8

中国版本图书馆 CIP 数据核字（2014）第 019448 号

书　　名：统计学(第2版)
著作责任者：阮红伟　主编
责 任 编 辑：刘　丽
标 准 书 号：ISBN 978-7-301-23854-7/C・0985
出 版 发 行：北京大学出版社
地　　址：北京市海淀区成府路 205 号　100871
网　　址：http://www.pup.cn　新浪官方微博:@北京大学出版社
电 子 信 箱：pup_6@163.com
电　　话：邮购部 62752015　发行部 62750672　编辑部 62750667　出版部 62754962
印　刷　者：
经　销　者：新华书店
787 毫米×1092 毫米　16 开本　17.25 印张　392 千字
2009 年 7 月第 1 版　2014 年 3 月第 2 版　2014 年 3 月第 1 次印刷
定　　价：35.00 元

前　　言

现在的社会是一个知识经济的社会，我们所处的时代是一个信息的时代、竞争的时代。统计是获取信息的手段，而且具有反馈信息、处理信息、咨询监督、支持决策的作用。统计学作为一门收集、整理和分析数据的方法论科学，在这个瞬息万变的现代社会将发挥更加巨大的作用。越是先进的社会，统计理论和统计方法的普及率、应用率越高。统计学的基础性、工具性决定了它在社会实践中应用的广泛性。能将普通基本的统计知识与平凡的工作岗位相结合，不断解决本部门、本岗位的客观实际问题，这就是一种创新精神和实践能力的体现。统计学的方法、思维方式和理念与时代紧密相连。学习统计学，掌握其工具和方法，培养统计思想和观念，不仅能为后续专业课的学习打下坚实的基础，更对未来人生观有重要的指导意义。

本书是在总结了编者多年的统计教学实践经验的基础上写成的，自 2009 年 7 月第 1 版出版以来，先后印刷多次，获得广大读者的接受和肯定。本次编写之前，编者通过问卷调查及座谈方式，对部分读者进行了调查，在广泛吸取读者意见的基础上，对第 1 版进行了修订。

本书按照统计工作的步骤书写，包括统计概论、统计调查、统计整理、综合指标、时间数列、统计指数、抽样推断、相关分析与回归分析等内容。本书紧密结合当前高等院校经济管理类人才需求和发展前景，以培养学生创新精神和提高学生实践能力为目的，强调实际操作训练，融“教、学、做”为一体。全书没有采用高深的理论推导和烦琐抽象的公式论证，内容既系统又简明易懂。本书中的案例及练习均取材于实际经济生活，很多来源于近年相关媒体发表的资料，内容通俗、生动，有趣味性和吸引力，同时又便于学生在实践中参照应用。

本书采用一体化设计，每章在列出学习该章的理论目标和能力目标后，提出日常经济生活中遇到的实际问题让学生亲自做一做、想一想，在理论构建和工具方法上承上启下、由浅入深，引出解决问题的该章相关理论。章节内容的讲授注重直观实用，避免过多的公式推导。每章后设置了该章小结，强化知识的逻辑关联点，归纳该章的关键概念和主题；随后的应用与拓展，紧密联系该章内容，贴切地解决社会经济中的现实问题，拓展学生的创新思维；最后的习题与实训，融入“情景”案例、“实用”项目，提升学生的技能水平和学以致用的成就感。本书内容的整体设计体现了“工学交替”“任务驱动”“项目导向”的理念。

本书的这次修订主要包括以下内容：一是根据当前就业市场的需求，着眼于“实用型”教育，增添了反映国家统计局最近颁布的新规范、新标准的内容；二是更新了教材案例，加大了教材案例化程度，使案例更密切地反映现实经济社会现状，增进其趣味性和时效性；三是增加了“习题与实训”的比例，更换其中内容，融入鲜活实用的实例；四是对第 1 版中不妥之处进行了修正。因此，本书具有实用性、先进性、创造性及趣味性的特点。

本书由青岛大学阮红伟副教授担任主编，由张丕景教授担任主审，王希兴副教授参加了部分章节的编写。全书结构体系及统编定稿由阮红伟完成。

编者在修订本书过程中，参考和借鉴了大量的国内外同类著作、文献以及报刊资料，值此出版之际，特向这些作者表示诚挚的感谢！同时，在此对为本次修订提出宝贵意见的教师和读者，以及对本书的出版给予了大力支持的北京大学出版社的相关工作人员，一并深表谢意！

由于编者水平有限，书中的不足和疏漏之处在所难免，恳请广大读者给予批评和指正，以便改进和完善。

阮红伟
ruanhwqdu@163.com
2014 年 2 月

目　　录



第1章

统计概论

理论目标

(1) 理解统计的含义、研究对象和特点。
(2) 掌握统计的基本概念及各概念之间的区别与联系。
(3) 了解统计研究的基本方法和一般过程。

能力目标

(1) 能定义"统计"，并能阐述统计在实践中宽广的应用领域。
(2) 能针对不同的研究问题选取恰当的计量尺度进行计量。

做一做

杰信公司目前正在对本公司年龄在30岁以下的青年员工进行专业理论培训。在公司本部工作的青年员工有32人，组成A班；在公司分部工作的青年员工有42人，组成B班。在公司统一组织的第一次考试中，A班平均成绩为81分，B班平均成绩为76分；在刚刚结束的第二次考试中，A班不及格人数为5人，而B班不及格人数为6人，同时，A班平均成绩为80.2分，B班平均成绩为79.8分，假如你是分管此项培训工作的管理人员，请你根据这些信息，评价A、B两班的学习效果。

想一想

根据给出的资料，单独看不及格人数不能说明问题，因为这是绝对数，仅侧重于反映量的多少和规模大小，难以反映现象本质问题，若将不及格人数与全班总人数相比得到不及格率，则可以看出B班成绩好于A班(A班不及格率为15.63%，B班不及格率为14.29%)；但根据第二次平均成绩，则B班不

如A班；如果再考虑到第一次考试平均成绩，又会发现B班是进步的而A班是退步的。看起来数字之间似乎出现了矛盾的解释，原因何在？又该如何评价两个班的学习效果呢？

实际上，科学、全面地评判A、B两班的学习效果时，我们可以这样说，A班虽然平均成绩高于B班，但此班不及格率较高而且平均成绩比上次退步，应当继续努力，争取更好的成绩；B班虽然不及格率较低而且平均成绩比上次进步，但从平均成绩看，成绩低于A班，所以应当奋起直追，争取与A班并驾齐驱。做出这种结论的依据，是统计学中关于总量指标、相对指标和平均指标的特点和性质的理论。如果没有基本的统计知识，对身边这样的生活例子我们则难以理解透彻，并做出科学决策。统计是人们认识社会很重要的一个工具，本章将从总体上阐述统计研究的一般问题。

1.1 统计的性质和特点

1.1.1 统计的含义

统计(statistics)一词在不同的场合有3种含义：统计工作、统计资料、统计学。

统计工作(statistical operation)是指具体从事的统计设计、资料搜集与整理、分析预测，以及提供各种统计资料的实践活动的总称。例如，前面讨论的计算反映A、B两班成绩的各种指标，并且进行分析评价的工作过程就是统计工作。

统计资料(statistical data)是指在统计工作过程中所获得的以统计数据表现的信息资料。例如，反映A、B两班成绩的各种指标就是统计资料。

统计学(statistics)是阐述正确指导统计活动科学原理和方法的学科体系。从广义上讲，统计学是包括自然科学和社会科学在内的统计科学理论的总和。本书专门阐述作为社会科学分支的统计学理论和方法，即社会经济统计学，主要论述对社会经济现象如何进行统计设计、统计调查和统计整理及分析统计资料的理论和方法。

统计一词的3种含义有着密切联系。统计工作与统计资料是工作过程与成果的关系，统计学与统计工作是理论与实践的关系。因此，统计一词是统计工作、统计资料、统计学的综合概括，是统计的过程与结果、理论与实践的辩证与统一。

统计作为搜集客观实际资料的社会实践活动已经有几千年的历史了。但是，“统计”用语的出现只有不到300年的时间。18世纪，德国政治学家G. 阿亨瓦尔(G. Achenwell)最早把“国势学”定名为“statis tika”，即统计学。此后，各国相继沿用“统计”这个词，并把它译成各国文字。1903年，我国学者钮永建、林卓南等翻译了日本横山雅男所著的《统计讲义录》一书，统计这个词才传到我国。1907年，彭祖植编写的《统计学》是我国最早的一本统计学书籍。从此，“统计”一词在我国就成了记述国家和社会状况的数量表现及数量关系的总称。

1.1.2 统计研究对象的特点

社会经济统计的研究对象是大量社会经济现象总体的数量方面，其特点可概括为以下3个：数量性、总体性和社会性。



1. 数量性

统计的研究对象并不是社会经济现象总体的各个方面，而是研究其数量方面，即以数据为依据具体说明社会经济现象总体的数量特征、数量关系和数量界限。例如，社会人口的数量及构成，社会财富的数量及分配，经济发展的规模和速度，人民群众物质文化水平现状及变化状况等。利用反映这些现象数量方面的各种统计数据，能说明社会经济现象发展状况、发展变化关系及预测未来，生动形象，具有较强的说服力。

统计研究对象的数量性，是统计区别于其他社会经济调查研究活动的根本特点。但是，必须指出，统计是在质与量的辩证统一中研究社会经济现象总体的数量方面的。社会经济统计不是“纯数量”的研究，这是统计与数学的重要区别。统计反映的数量具有具体性，而数学反映的数量具有抽象性。例如，反映社会经济现象特征的统计数据，都是明确规定了具体的时间、地点条件下的数量。2013 年(时间)我国(地点)初步核算的国内生产总值为 568 845 亿元(国家统计局，2014 - 02 - 24)，这个数量就是 2013 年我国的国内生产总值数量，而不是其他国家、其他时间条件下的数量。如果抽掉具体的内容，不是在一定时间、地点和条件下进行研究，那就不能说明任何问题，也就不能称其为统计，其数字也就不是统计数字。

统计的数量性与会计学反映的数量也有区别。会计学主要研究现象的价值量，而统计学不但研究价值量，还要研究实物量和劳动量；会计核算主要描述数量表现，统计研究不但描述数量表现，还要研究数量关系和数量界限。另外，统计与其他学科也有关系，统计可以帮助其他学科探索其学科内在的数量规律性，而对这种规律性解释的深入研究则由各学科完成。

2. 总体性

统计的研究对象不是个体现象的数量方面，而是社会经济现象总体的数量方面。统计的研究结果都是描述总体特征的。但是总体是由个体所构成的，要认识社会经济现象总体，必须从个体入手进行调查研究。“研究个体”是过程和手段，“反映总体”是结果和目的。例如，要研究某班全体同学的某门课程的平均成绩，必须搜集每位同学的成绩，然后进行汇总计算，所得平均分数并非指某位同学的成绩，而是代表全班总体的水平。

3. 社会性

统计的研究对象是社会经济现象。它包括人类活动的各种条件，如社会条件、自然条件；包括人类各种活动的过程和结果，如生产活动、交换活动和消费活动等。因此，社会经济统计在研究社会经济现象时，不是孤立地进行的，而是要联系有关社会经济现象进行全面具体的分析，同时也要联系有关自然现象与技术因素等方面进行研究，即具有明显的社会性。这样才能说明现象变化的原因与过程，科学认识社会经济现象。

1.1.3 统计的分类

统计学的内容十分丰富，研究和应用的领域非常广泛。从统计教育的角度讲，统计学

的分类大致有以下两种。

1. 理论统计学与应用统计学

理论统计学(theoretical statistics)主要是指统计学中关于数据的搜集、整理和分析的最基本的原理、原则和方法，一般既适用于社会经济现象数量特征的观察和分析，也适用于自然现象和科学实验数据的分析研究，它是统计学应用于各种领域的理论基础。从这个角度讲，统计学是一门通用的方法论科学。

应用统计学(applied statistics)是运用于某一特定领域的统计理论与方法。统计学是一门数据科学。由于在自然科学、社会科学的所有研究和实际工作中，都要通过数据分析问题和解决问题，统计方法的应用就自然而然地扩展到几乎所有的研究领域，形成了各种应用统计学。例如，统计方法在医药领域的应用就形成了医药统计学，统计方法在人口领域的应用就形成了人口统计学，统计方法在工业领域的应用就形成了工业统计学，统计方法在经济、贸易、管理、商务等领域的应用就形成了相应的经济统计学、贸易统计学、管理统计学、商务统计学等。应用统计学的不同分支所应用的基本原理和方法是一样的，但由于每个领域都有其特殊性，所以统计方法在不同领域的应用具有不同的特点。

2. 描述统计学与推断统计学

描述统计学(descriptive statistics)研究如何对客观现象的数量特征进行观察、搜集数据进行计量并给予概括和表述。具体地说，描述统计学的主要内容包括：确定所要研究的数量特征及其计量层次；设计用来说明现象的数量特征的统计指标；搜集数据并对数据进行整理；计算统计指标并用图表表示。例如，描述某班级总人数，某次考试平均成绩；描述"十一"黄金周期间，北京故宫旅游景点总收入；描述2005—2012年，海尔集团平均投资回报率等。

推断统计学(inferential statistics)研究如何从总体中抽取部分样本，并根据样本数据推断总体的数量特征。例如，公司在准备投入新产品时，通常需要估计相关市场上的消费者偏好。这可以通过对随机抽取的若干家庭进行市场调查实现，调查的结果可以作为总体消费者偏好的估计。当买家从供应商那里收到大批装载的货物，买家肯定希望确保货物的质量达到合同要求。如果进行全面的质量检验，将花费大量成本，甚至某些破坏性检查将会使产品报废，而通过随机抽取一定样本进行检测以推断出总体的质量，则可使问题得以解决。为了保证这种推断的精确度和可靠度，推断统计学需要研究样本的抽取方式、样本的大小、样本的分布、样本估计量的选择、对总体特征进行估计或推断的方法、误差的计算和控制等问题。

描述统计学和推断统计学，一方面反映了统计发展的前后两个阶段；另一方面也反映了统计方法研究和探索事物内在规律性的先后两个过程。为达到统计研究的目的，如果所搜集到的是总体数据(普查)，那么，采用描述统计方法可直接达到研究目的；如果所搜集到的只是所研究总体的一部分数据，即样本数据，则采用描述统计方法得出样本的数量特征后，还必须用推断统计方法根据样本整理出的信息对总体做出科学的推断。可见，描述统计学是整个统计学的基础，而推断统计学是现代统计学的核心。



1.2 统计工作过程与研究方法

1.2.1 统计工作过程

统计是认识社会的工具。统计对社会经济现象的研究过程，也就是对整个社会经济现象的认识过程。统计的认识活动与人类其他的认识活动一样，是一个由浅入深、由表及里、不断深化、永无止境的过程。随着客观事物的不断发展变化，统计认识活动需要不断进行。一般而言，一个完整的统计工作过程可以分为4个主要阶段，即统计设计、统计调查、统计整理和统计分析。

1. 统计设计

统计设计是统计工作的第一个阶段。它是根据统计研究对象的性质特点和统计研究目的，事先对统计工作的各方面和各环节所进行的通盘考虑和安排。

统计设计的主要包括以下内容。

(1) 统计指标和指标体系的设计。这是统计设计的核心内容。统计设计要根据统计的任务、目的及研究对象的特点，选择能反映现象本质特征的指标组成指标体系，同时，还要考虑指标间的相互联系，明确指标口径范围、计算方法和重要的分组等。

(2) 搜集整理资料方法的设计。根据统计的目的和任务确定适当的搜集资料方法，如普查、抽样调查、重点调查等。统计整理同样也有多种方法，需要预先选择好。

(3) 统计工作各部门和各阶段的各种保证条件的要求。统计各部门的协调、统计各阶段的联系、统计活动需要的人员和经费等，都需要预先考虑安排好。

(4) 具体实施方案的设计。在考虑了上述3个方面之后，要求具体安排各个环节，提出日程表和工作进度，以便监督实施。

统计设计的结果表现为各种设计文案，主要包括：各种标准、制度、规定、方案和方法等。例如，统计分类标准和目录、统计指标体系、统计报表制度、统计调查方案、统计整理和汇总方案等。

统计设计在统计工作中起着重要作用。统计研究对象是社会经济现象总体，往往涉及面广、工作量大，投入的人力、物力、财力较多，对经济活动影响较大，这就要求统计工作要高度集中统一。无论是统计总体范围、统计指标口径和计算方法，还是统计分类和分组标准，都必须统一。因此，只有事先进行统计设计，才能做到统一认识、统一步骤、统一行动，使统计工作有秩序地协调进行，保证统计工作的质量。

2. 统计调查

统计调查是搜集统计资料的工作过程。它的任务是根据统计设计的要求，采用科学方法，针对社会经济现象有计划、有组织地搜集统计资料。

统计调查是统计的基础工作，是统计活动定量认识的起点，非常重要。统计调查阶段所搜集的资料是否准确、及时、全面、系统，将直接关系到统计整理的好坏，影响统计分

析的结论，决定统计工作的质量。

统计调查的结果表现为各种调查表、登记表等原始数据的反映。

3. 统计整理

统计整理是根据统计研究的目的，将统计调查所搜集到的原始资料进行科学的分类和汇总，使之系统化、条理化，为统计分析提供能描述现象总体数量综合特征的资料的工作过程。

统计整理是统计工作的中间环节。统计整理能帮助统计工作者由对社会经济现象的个体认识过渡到对总体的认识，由感性认识过渡到理性认识。统计整理是统计调查阶段的深入和继续，又是统计分析阶段的基础和前提，起着承上启下的作用。

统计整理的结果表现为各种整理表、统计图等。

4. 统计分析

统计分析是对经过加工整理的统计资料进行分析研究，采用各种统计分析方法，计算各种统计分析指标，目的是认识和揭示现象的本质和规律，得出科学结论，进而进行预测或作为决策依据的工作过程。

统计分析是统计工作的最后阶段，属于认识的理性阶段，是统计研究的决定性环节。有了科学的统计分析结论，才能充分发挥统计的信息、咨询和监督职能。

统计工作过程的 4 个阶段是相互联系、不可分割、依序进行的。在实际工作中，只有做好每一阶段的工作，才能保证整个统计工作高质、高效地完成。

1.2.2 统计研究方法

在长期的统计实践活动中，人们根据统计研究对象的特点和研究目的的需要，总结和创造出了一系列科学的统计研究方法，如大量观察法、统计分组法、综合指标法、动态分析法、指数分析法、抽样推断法、相关分析法、统计预测法等。这一系列方法的有机结合，构成了统计特有的研究方法体系，其中，大量观察法、统计分组法、综合指标法是最基本、最主要的研究方法。准确把握这些方法的基本思想和精神实质，对于做好统计工作具有十分重要的意义。

1. 大量观察法

大量观察法就是对现象总体中的全部或足够多数的单位进行观察，并加以综合分析的方法。统计学研究的是社会总体而不是个别的社会现象，由于社会现象的复杂性和总体性，必须对总体进行大量观察和分析，研究其内在联系，方能反映社会现象的规律。例如，人口现象中的男女比例问题。就单独一个家庭来观察，新生婴儿的性别可能是男性，也可能是女性，从表面上看，新生婴儿的性别比例似乎没有什么规律可循。但如果对大量家庭的新生婴儿进行观察，就会发现新生婴儿中男孩略多于女孩，大致为每出生 100 个女孩相应地就有 107 个男孩出生。这个性别比例 107∶100 就是新生婴儿性别比的数量规律，古今中外这一比例都大致相同，这是由人类自然发展的内在规律所决定的。尽管从新生婴



儿看，男性婴儿略多于女性，似乎并不平衡，但由于男性的死亡率高于女性，到中年时，男女人数就大体相当了。进入中老年后，男性死亡率仍然高于女性，导致男性的平均预期寿命比女性短，老年男性反而少于老年女性。只有对整个社会的所有家庭或足够多数的家庭成员进行调查，才能准确揭示人口现象中男女比例的一般特征和规律。大量观察法是社会经济统计学的基本观察方法。

2. 统计分组法

统计分组法就是根据现象的特点和统计研究的目的，将现象按不同类型或不同性质划分成若干个部分的统计方法。社会经济现象总体是由许多具有某种同质性的单位组成的群体，但由于在不同总体范围内的单位之间具有一定差别，因此有必要进行统计分组，以区分社会经济现象的不同类型和形态。例如，在社会人口这一统计总体中，就存在着年龄、学历、职业和收入上的种种差别。统计分组把总体内不同性质的单位区别开来，使性质相同的单位归在一个组内，可以区分现象的类型，反映总体的结构，揭示现象间的依存关系，从而使得对统计总体的数量表现、数量关系和数量界限的研究更加深入和透彻。

3. 综合指标法

综合指标法是指在统计研究中运用总量指标、相对指标、平均指标等各种综合指标对大量社会经济现象的数量方面进行综合分析，概括地表明社会经济现象总体数量特征的方法。统计作为认识社会的工具，必须对社会现象进行全面、科学的描述。总量指标是数量指标，侧重反映事物“量”的特征；相对指标和平均指标是质量指标，侧重反映事物“质”的特征。在统计研究中，只有综合运用这些指标，才能做到从质与量两方面反映问题，多角度、全方位、科学、公正地反映现象总体的特征。

1.3 统计学的基本概念

统计学中有许多概念，其中有几个概念是我们经常要用到的基本概念。下面对这些基本概念进行介绍。

1.3.1 统计总体与总体单位

1. 统计总体

统计总体简称总体(population)，是根据统计研究目的确定的所研究对象的全体。它是由客观存在的、具有某种共同性质的许多个体组成的。例如，全国总人口构成的整体，上半年某公司销售的彩电构成的总体，2013 年 4 月份某地发生的交通事故构成的总体等。

总体具有以下 3 个方面的特征。

(1) 同质性。即构成总体的每个个体必须具有某一方面的共性。

(2) 大量性。即总体是由许多个体所组成的，而不能只是个别或少数单位。这是因为统计研究的目的是要揭示现象的规律性，而社会现象的规律性只有在大量现象的综合汇总中才能显示出来。个别单位的现象有很大的偶然性，而总体相对稳定，表现出共同性的倾向。

(3) 差异性。即总体的各个个体之间，除了必须在某一方面有共性之外，在其他方面还应存在差异，这些差异是统计研究的基础。

统计总体的范围可大可小。例如，要进行全国范围的人口研究，全国总人口是统计总体，总体范围较大；而要研究某班级学生的身体状况，则全班同学是统计总体，总体范围较小。由于统计研究的目的和任务不同，统计总体既可以由人(如一名学生)、物(如一台彩电)、事(如一起交通事故)构成，也可以由企事业单位(如一所学校)构成。统计总体可以分为无限总体和有限总体两类。无限总体是指包括的个体数很多，以致无法计量的总体。例如，宇宙中星球的个数、海洋鱼类等都是无限总体。社会经济统计中，经常把企业自动化生产线上大量的、连续不断生产出来的小产品或零件看成无限总体。有限总体指由有限个个体构成的总体。例如，全国高等院校组成的总体，某城市全体居民组成的总体。社会经济统计中，大多数都是有限总体，既可以采用全面调查，也可以采用非全面调查。而对于无限总体，只能采用非全面调查方法来推断总体情况。

从总体中抽取一部分个体构成的集合，称为样本(sample)。构成样本的个体的数目，称为样本量。例如，从一批袋装奶粉中随机抽取100袋，这100袋奶粉就构成了一个样本，100袋是样本量。抽取样本的目的，就是要根据这100袋奶粉的检测指标推断这一批袋装奶粉的质量。

样本的单位数相对于总体来说，虽然只是很少的一部分，但样本是从总体中产生并代表总体的特征。因此，样本所取自的总体又称为母体，样本又称为子样。

样本是统计学中非常重要的概念。样本既然是从总体中以某种方式抽取的，它们就具有与总体同质的数量特征。但由于每次抽取样本时客观条件可能不同，因此，即使对同一个总体按照同一样本量抽取样本，得到的样本也是多种多样的。所以，样本具有随机性。

2. 总体单位

组成总体的每个个体称为总体单位。由于统计总体可以由人、物、事构成，所以总体单位既可以是人，也可以是物或事。例如，在由全国总人口构成的总体中，每一个人就是总体单位；在上半年某公司销售的彩电构成的总体中，每台彩电就是总体单位；在2013年4月份某地发生的交通事故构成的总体中，每一起交通事故就是总体单位。

很显然，统计总体是由总体单位组成的，但统计总体与总体单位的概念又是相对而言的，随着研究目的和总体范围不同可以相互变化。同一研究对象，在一种情况下为总体，但在另一种情况下又变成了总体单位。例如，要研究广州市所属各辖区人口数时，广州市为统计总体，每个辖区为总体单位；而要研究全国各城市人口数时，则全国所有城市是统计总体，广州市又成了总体单位。



1.3.2 标志与指标

1. 标志

标志(mark)是反映总体单位的属性和特征的名称。例如，某班级学生构成一个统计总体，每一位学生是这个总体的总体单位，反映学生的各种特征的名称如性别、籍贯、年龄、身高、学习成绩等都称为总体单位的标志。

标志可分为品质标志和数量标志。品质标志是指反映总体单位质的特征的名称，只能用文字表示，不能用数值表示。例如，学生的性别、籍贯、民族等，都属于品质标志。数量标志是指反映总体单位量的特征的名称，一般用数值表示。例如，学生的年龄、身高、学习成绩等，都属于数量标志。

标志还可分为不变标志和可变标志。不变标志是指所有的总体单位共同具有的特征。例如，在“高等院校女生总体”中，每一位女生是总体单位，性别是反映总体单位特征的标志，它在女生总体中不发生变化，即大家都是女性，此时，性别这个标志就是不变标志。可变标志是指在总体各单位之间存在差异的标志。例如，在“高等院校女生总体”中，身高、体重、年龄等标志在每位女同学之间都存在差异，则它们都是可变标志。不论是数量标志还是品质标志，都有可能是可变标志。统计学上把标志在各总体单位上所表现出来的差别叫做变异，所以可变标志也称为变异标志。不变标志的存在保证了统计总体的同质性，是构成统计总体的必要条件和确定总体范围的标准；可变标志的存在保证了统计总体的差异性，是进行统计研究的兴趣和目的所在。

标志和标志表现是两个不同的概念。尽管总体各单位都具有共同的标志，但每个单位对该标志的具体表现却可能不同，而总体各单位在特定时间、地点条件下的具体表现正是统计最关心的问题。品质标志与数量标志表现各不相同。品质标志表现只能用文字、语言来描述。例如，性别是品质标志，男性或女性则是标志的具体表现；籍贯也是品质标志，其标志表现具体为山东省、江苏省、广东省等。数量标志表现是用数值来表示的。例如，身高是数量标志，165cm、170cm、175cm 等是数量标志表现；学习成绩是数量标志，具体分数 80 分、90 分等是数量标志表现。数量标志表现又称为标志值。

一个标志当它既是数量标志又是可变标志时，我们称其为变量。在“高等院校女生总体”中，身高、体重、年龄等标志都是可变的数量标志，因而都可以称为变量。某一变量在各总体单位的具体表现称为变量值。例如，身高是变量，160cm、170cm、180cm 等都是变量值。可见，一个变量有若干个变量值。

变量(variable)根据其变量值的取值情况可分为连续型变量(continuous variable)和离散型变量(discrete variable)。连续型变量是指相邻的两个变量值之间可以有无限多个数值，变量值是连续不断的，取整数或取小数都具有经济含义。例如，人的身高、体重，某地的国内生产总值，固定资产投资额等都属于连续型变量。离散型变量是指两个相邻的变量之间没有小数，变量值只能用整数表示，当取小数时，变量就失去了经济含义。例如，各公司的员工数，家庭拥有的移动电话数、彩电数等，其取值只能用整数表示，因而均属于离散型变量。

2. 指标

指标是统计指标(statistical indicator)的简称，是指用于反映统计总体数量特征的概念或范畴。它是统计学中一个重要概念，可以反映总体的规律性。一项完整的统计指标主要由4部分构成，即指标名称、指标数值、指标单位和指标条件。例如，“我国2013年国内生产总值568 845亿元”，其中，“国内生产总值”是指标名称；“568 845”是指标数值；“亿元”是指标单位；“我国2013年”是指标条件，即指标的空间及时间范围。在社会经济生活中的统计设计阶段，时常把调查项目名称叫做统计指标，而在调查汇总之后又把指标数值叫做统计指标，这都不影响统计指标构成的完整性。统计指标具有对统计总体特征定性认识和定量认识的双重作用。

统计指标可分为数量指标和质量指标。数量指标是反映总体绝对数量多少的总量指标。例如，全国人口总数、高等院校数、国内生产总值、销售收入、出口贸易总额等，这些指标无论是总体单位总量，还是总体各单位标志总量，都反映了现象的总规模、总水平和工作总量，所以，数量指标也称为总量指标，用绝对数表示。质量指标是反映现象相对水平和工作质量的统计指标。例如，人均国内生产总值、职工平均工资、产品合格率、学生出勤率等。质量指标是总量指标的派生指标，反映现象之间的内在联系和对比关系，用相对数或平均数表示。

反映总体特征的指标叫参数(parametric)，反映样本特征的指标叫做统计量(statistic)。参数与统计量的划分只有在推断统计学中才重要。本章前面提到的推断统计学，实际上就是根据样本的统计量来推断总体参数的过程。

指标和标志是既有联系又有区别的两个概念。两者的区别是：指标是反映总体特征的，而标志则是说明总体单位的；指标(不论是数量指标还是质量指标)都能用数量来反映，而标志中只有数量标志才能用数值来表示(品质标志不用数值表示)。两者的联系是：标志是指标的基础，没有标志和标志表现，就没有指标；指标和标志的确定并非一成不变，当总体和总体单位随统计研究目的发生变化时，指标和标志也必然随之发生相应的变化。

单个统计指标只反映总体某一个个数量特征，说明现象某一个侧面情况。然而，客观现象是错综复杂、相互联系的，要科学、全面地反映社会经济现象的数量特征，描述现象发展的全过程，需要将一系列有联系的统计指标有机地结合起来进行分析研究。由一系列相互联系的统计指标所组成的有机整体，称为统计指标体系。例如，为了反映公司生产经营状况，只设立利润这一项指标是不够的，应该考虑公司的投入、投入与产出之比，所以，应该设立产量、增加值、职工人数、工资总额、利润、产值、劳动生产率、产值利税率、资金成本利润率等一系列指标构成的指标体系，才能反映公司经营全貌。社会经济统计指标体系大体上可分为两大类，即基本统计指标体系和专题统计指标体系。基本统计指标体系是反映国民经济和社会发展及其各个组成部分的基本情况的指标体系，它包括反映整个国民经济和社会发展的统计指标体系，各地区和各部门的统计指标体系，以及基层统计指标体系。专题统计指标体系是针对某一个经济或社会问题而制定的专项指标体系，如小康生活水平指标体系、政府绩效评估体系等。



1.4 数据的计量尺度

统计的研究对象是社会经济现象的数量方面，统计数据是统计工作的基础，没有统计数据，一切统计方法就变成了“无米之炊”。统计数据的表现丰富多彩。例如，有代表产品成本的数据，有代表货物运输重量的数据，有代表排位名次的数据，有代表大型超市地理方位的数据，等等。这些数据所代表的含义不同，肯定不能采用相同的方法进行计量，所以，在进行统计工作搜集数据之前，就涉及数据的计量尺度问题。

统计数据可以是定性型的，如考生的性别、产品的质量等级等，它们只对客观现象的属性加以分类；统计数据也可以是定量型的，如农民的年收入、公司的利润等，它们可对客观现象用较精确的数字或数值加以描述。按照统计的一般分类，统计数据可细分为4种计量尺度：定类尺度(nominal level)、定序尺度(ordinal level)、定距尺度(interval level)和定比尺度(ratio level)。其中，定类尺度和定序尺度是定性数据(qualitative data)；定距尺度和定比尺度是定量数据(quantitative data)。

1.4.1 定类尺度

定类尺度是一种精确水平最低、最粗略的计量尺度。它仅仅是一种标志，用以区分变量的不同值，没有次序关系。例如，在统计问卷调查中，特别是人口统计中，经常调查人口的性别、职业的分类等。为了便于统计整理汇总，特别是为便于计算机的识别，通常给每类或每组一个数码，如数码1表示男性，数码2表示女性，或者反过来编码，数码1表示女性，数码2表示男性。也可以用其他任意两个不同的数码代表类别。这些数码都仅仅是一个符号，不能代表次序排列，也不能用来进行运算。例如，在这里 $1+2\neq3$。在社会经济现象中，如宗教、种族、地理方位、出生地、社会保险号码、电话号码、身份证号码等都可以用定类尺度计量。适用于分析定类数据的统计技术是有限的，然而，统计的一些更宽广的应用，如 χ^2 统计能够利用定类数据产生有用的信息。

1.4.2 定序尺度

定序尺度比定类尺度的精确性要高一些，定序数据表现为各类别或各组之间有一定的顺序，数据可以排序，而且可以比较大小。例如，学生的考试成绩可分为：优=1，良=2，中=3，及格=4，不及格=5。当然，也可以倒过来用5～1表示不同等级，信息一点也没有损失。各编码的序值，代表了考试成绩的高低差异。尽管这个差异不能准确描述差别的大小，但可以确定其顺序，即可对定序数据做不等式运算。例如，对考试成绩而言，如果已知“优>良”，“良>中”，则肯定有“优>中”。

在社会经济现象中，《财富》杂志对世界500强公司的排名，针对某服务项目按照“很好、较好、一般、较差、很差”的级别进行评价，观众对电视栏目“强烈推荐、推荐、不推荐”的评级等都需要用到定序尺度来计量。

定序数据提供了有关相对比较方面的信息，但没有提供差别程度，它与定类数据一样也不能用来进行计算。定类数据和定序数据又称为定性数据。

1.4.3 定距尺度

定距尺度是比定序尺度高一层次的定量尺度。它不仅能将总体各单位分类，并使各类型具有实质性意义的排序，而且能以确切的数值反映各单位之间间距的大小，所以它的精确性比定类尺度前进了一大步。它是定比尺度的基础，在统计数据中居于重要的地位。例如，温度21℃、22℃、25℃等能够按照温度高低进行排序，同时，能准确测定两个温度之间的差距，即21℃与22℃相差1℃、22℃与25℃相差3℃。

定距尺度可以进行加法运算，也可以进行减法运算，它们都有实际意义。但是，定距尺度不能进行乘除运算。因为，在等级序列中没有固定的、有确定意义的“0”位。0仅仅是刻度上的另一个点，而不意味着现象不存在。例如，0℃并不是最低的、起点温度，因此不能说20℃是10℃的2倍。实际上，将这两个摄氏温度换算为热力学温度后，这两个温度并不是2倍的关系。

在社会经济现象中，城镇登记失业率、就业率、投资回报率等都是定距数据。

1.4.4 定比尺度

定比尺度是数据最高级的定量尺度。定比尺度具有定距尺度的特点，但定比尺度有绝对的0，并且两个数据之比是有意义的。“绝对的0”的意思是说，0是固定的，而且“0”表示“没有”或“无”。例如，高度、重量、容积、产量、利润等均为定比数据。在这里，“0”表示该事物不存在，而且，两数据的比率具有意义。我们可以说，重量180kg是90kg的2倍，换句话说，两个重量之比是180：90。

在社会经济现象中，旅客乘车到达目的地的路程、产品生产周期、职工人数、各地区国内生产总值、公司的利润等都是定比数据。定比尺度在统计实践中应用非常广泛。

定比尺度与定距尺度的差别在于，定比尺度具有一个自然的零始点，数据之间的比率是有意义的。但两者通常都是在社会各领域中经过精密测算而得到的。定距数据和定比数据又称为定量数据。

1.4.5 4种计量尺度的比较

定类尺度、定序尺度、定距尺度和定比尺度的关系如图1.1所示。

从最中心的定类尺度到最外围的定比尺度，它们是按照由较低的或较粗略的计量到较高的或较精密的计量排序的。一般来说，较精密的计量尺度可以转化为较粗略的计量尺度，亦即定量数据可以转化为定性数据。例如，定距尺度和定比尺度是可变成定类尺度和定序尺度的。但较粗略的尺度不能转变为较精密的计量尺度。而且，不同级别的计量尺度，应用范围不同。一般来说，数据的等级越高，应用范围越广泛；等级越低，应用范围越受限制。

统计尺度分4个不同的层次，只是针对客观事物量化程度和运算功能来说的，而不是指统计研究本身的高低之分。统计技术可分为参数统计和非参数统计两种。参数统计要求数据是定距数据或定比数据，如果是定类数据或定序数据，则应采用非参数统计。非参数



统计也可以用来分析定距数据或定比数据。本书内容主要是参数统计，因此，定量数据是我们研究的重点。

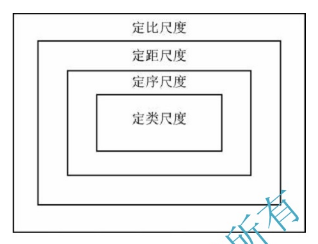


图 1.1 4 种计量尺度关系

本章小结

本章共分 4 节，前两节主要从总体上阐述了研究统计的一般问题，内容包括统计的含义、统计的研究对象、统计学的工作过程和研究方法等；第 3 节讲述统计学的基本范畴，第 4 节介绍统计计量尺度。

统计一词有统计工作、统计资料、统计学 3 种含义，它的研究对象是大量社会经济现象总体的数量方面，其特点归纳起来可概括为数量性、总体性和社会性。一个完整的统计工作过程可以分为 4 个主要阶段，即统计设计、统计调查、统计整理和统计分析。统计研究从社会经济现象质量与数量的辩证统一中研究现象的数量方面，运用各种研究方法，其中，大量观察法、统计分组法、综合指标法是最基本、最主要的方法。

统计总体是根据统计研究目的确定的所研究对象的全体。它由客观存在的、具有某种共同性质的许多个体组成。统计总体必须同时具备同质性、大量性、差异性。统计总体可大可小，可以是有限总体，也可以是无限总体。组成总体的每个个体称为总体单位，它可以是人、物、事件等。

样本是从总体中抽取一部分个体构成的集合。构成样本的个体的数目，称为样本量。样本具有随机性。

标志是反映总体单位的属性和特征的名称。它可分为品质标志和数量标志、不变标志和可变标志。可变的数量标志称为变量，其具体表现的数值称为变量值。按变量值取值的不同，变量分为连续型变量和离散型变量。

指标是指用以反映统计总体数量特征的概念或范畴，它可分为数量指标和质量指标。反映总体的指标叫参数，反映样本的指标叫统计量。一系列相互联系的统计指标所组成的有机整体则称为统计指标体系。

统计搜集数据要涉及计量尺度的概念。4 种计量尺度是：定类尺度、定序尺度、定距尺度和定比尺度。定类尺度和定序尺度是定性数据，定距尺度和定比尺度是定量数据。

应用与拓展

2012年中秋节国庆节假日旅游统计报告

参照国家旅游局、国家统计局联合制定的《"黄金周"旅游信息统计调查制度》的要求，全国31个省、自治区、直辖市分别对辖区内中秋节国庆节假日的接待规模和效益进行了统计调查。经国家旅游局、国家统计局汇总，结果如下。

(1) 在今年中秋节国庆节8天假日期间，全国共接待游客4.25亿人次，比2011年"十一"黄金周增长40.9%(按可比口径，同比增长23.3%)；实现旅游收入2 105亿元，比2011年"十一"黄金周增长44.4%(按可比口径，同比增长26.3%)；游客人均花费支出495元。

(2) 中秋节国庆节假日期间，在全国接待的4.25亿人次旅游者中，过夜游客(仅限于住在宾馆饭店和旅馆招待所)为9 694万人次，比2011年"十一"黄金周增长36.0%(按可比口径，同比增长19.0%)；一日游游客为3.28亿人次，比2011年"十一"黄金周增长42.4%(按可比口径，同比增长24.6%)。在所实现的2 105亿元旅游收入中，民航客运收入81.6亿元；铁路客运收入33.2亿元；39个重点旅游城市实现旅游收入786亿元；其他旅游城市和景区实现旅游收入1 204亿元。

(3) 纳入全国假日旅游统计预报体系的北京、天津、承德、秦皇岛、沈阳、大连、长春、吉林、哈尔滨、上海、南京、无锡、苏州、杭州、宁波、黄山、厦门、南昌、瑞金、青岛、洛阳、武汉、长沙、张家界、韶山、广州、深圳、桂林、海口、三亚、重庆、成都、广安、贵阳、遵义、昆明、西安、延安、银川39个重点旅游城市，中秋节国庆节假日期间共接待游客1.39亿人次，其中过夜游客(口径同上)为2 856万人次，一日游游客为1.10亿人次。

(4) 纳入全国假日旅游统计预报体系的119个旅游景区(点)，中秋节国庆节假日期间共接待3 425万人次，门票收入17.7亿元。

(资料来源：http://www.cnta.gov.cn/html/2012-10/2012-10-9-15-13-73191.html.)



习题与实训

一、单项选择题

1. 统计研究最基本、最主要的研究方法有(　　)。

A. 问卷设计法、统计分组法、综合分析法

B. 抽样调查法、统计图表法、统计预测法

C. 大量观察法、统计分组法、综合指标法

D. 数量调查法、抽样推断法、统计分析法

2. 5位同学英语考级成绩分别为60分、81分、64分、87分、70分，则成绩是(　　)。

A. 品质标志　　B. 数量指标　　C. 变量值　　D. 数量标志

3. 对中国女排队员的身高进行度量采用的计量尺度最恰当的是(　　)。

A. 定类尺度　　B. 定序尺度　　C. 定距尺度　　D. 定比尺度

4. 某学院2012级全部大学生的平均年龄为19.76岁，这是(　　)。

A. 数量标志　　B. 数量指标　　C. 品质标志　　D. 质量指标



5. 以下为数量指标的是(　　)。

A. 城镇居民平均工资　　B. 集装箱吞吐总量

C. 每百元 GDP 能耗　　D. 人口密度

6. 将公司 700 名员工的工资额加起来除以 700，这是(　　)。

A. 对 700 个标志求平均数　　B. 对 700 个变量求平均数

C. 对 700 个变量值求平均数　　D. 对 700 个指标求平均数

7. 下面属于连续变量的是(　　)。

A. 家庭拥有的彩电数　　B. 学生每周上网次数

C. 居民每户人口数　　D. 职工每月工资收入

8. 在 4 种计量尺度中，(　　)。

A. 定序尺度可以转化为定距尺度

B. 定性数据可以转化为定量数据

C. 定比尺度是最高级计量尺度

D. 定类尺度是最高级计量尺度

9. 指标和标志两者的关系是(　　)。

A. 两者都是反映现象总体特征的

B. 两者均只能用数字表示

C. 标志是由指标汇总得到的

D. 两者概念可以相互转化

10. 下面的指标数值属于统计量的是(　　)。

A. 考察上一年出品的每一部商业电影的电脑记录后发现，平均放映时间是 152min

B. 调查某班全部学生，得到平均成绩为 81 分

C. 从出口的袋装茶叶中随机抽查 100 袋，测得平均重量为 454.6g

D. 调查某医院全体医生，测得春节期间平均每人休息时间为 3.7 天

二、多项选择题

1. 要了解某地区全部成年人口的就业情况，那么(　　)。

A. 全部成年人是研究的总体

B. 成年人口总数是统计指标

C. 成年人口就业率是统计标志

D. “职业”是每个人的特征，是数量指标

E. 某人的职业是医生，这里的“医生”是标志表现

2. 下列指标中是质量指标的有(　　)。

A. 人均 GDP　　B. 环保志愿者人数

C. 物价指数　　D. 城镇登记失业率

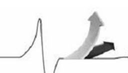
E. 第三产业增加值

3. 下列变量中是连续变量的有(　　)。

A. 公司利润　　B. 网络用户数

C. 存款余额　　D. 机器设备价值额

E. 学生人数



4. 要了解 100 个工业企业的生产情况，则统计指标是(　　)。

A. 100 个工业企业的工业增加值

B. 每一个工人的月工资

C. 某一个工业企业的工资总额

D. 全部工业企业的劳动生产率

E. 100 个工业企业的平均增加值

5. 总体与总体单位之间的关系有(　　)。

A. 总体与总体单位的概念是可以互换的

B. 总体可以转化为总体单位

C. 总体单位可以转化为总体

D. 总体与总体单位是固定不变的

E. 只能是总体转化为总体单位

6. 某高校在校学生 12 050 人，从中随机抽取 120 人进行调查，发现戴眼镜的有 71 人。在这里(　　)。



A. 在校学生 12 050 人是参数

B. 随机抽取的学生数 120 人是统计量

C. 戴眼镜人数 71 人是统计量

D. 3 个数值都是统计指标

E. 3 个数值都是数量指标

7. 在对“上月发生的交通事故”进行研究时，(　　)。

A. 每一起交通事故是总体单位

B. 每一起交通事故造成的经济损失金额是统计指标

C. 每一起交通事故的性质是品质标志

D. 每一起交通事故发生的具体地点是品质标志表现

E. 每一起交通事故都是统计总体

8. 下列各项中，属于统计指标的有(　　)。

A. 某地区出生人口总数

B. 某同学该学期的平均成绩

C. 我国体育代表团在北京奥运会上共获得 51 枚金牌

D. 2013 年上半年全国在岗职工人均工资

E. 2012 年年末全国农民工总量达到 26 261 万人

9. 在调查某城市商业网点时，(　　)。

A. 所有的商业网点总数是统计总体

B. 所有的商业网点总数是数量指标

C. 每一个商业网点是总体单位

D. 每一个商业网点的营业收入是数量指标

E. 平均每一个商业网点的营业收入是质量指标



10. 年终对员工进行质量考核，按照优秀、良好、合格、基本合格、不合格进行评价，若对数据进行计量应(　　)。

A. 采用定类尺度计量

B. 采用定性数据

C. 不能采用定距尺度或定比尺度计量

D. 采用定序尺度计量

E. 采用定量数据

三、判断题

1. 指标与标志是两个不同的概念，两者没有任何联系。(　　)

2. 办公室有5位职员，年龄分别是26岁、39岁、28岁、37岁、44岁，这些年龄是5个数量标志或5个变量。(　　)

3. 男性是品质标志。(　　)

4. 学生的平均身高是数量标志。(　　)

5. 人口普查中，全国总人口数是统计总体。(　　)

6. 数量标志可以用数值表示，质量指标不能用数值表示。(　　)

7. 数量指标的表现形式是绝对数，而质量指标的表现形式是相对数和平均数。(　　)

8. 只有对数量标志的标志表现进行汇总才能形成统计指标。(　　)

9. 统计工作与统计资料是工作过程与成果的关系，统计学与统计工作是理论与实践的关系。(　　)

10. 学生每周上网次数和时间均是离散型变量。(　　)

四、填空题

1. 统计一词在不同的场合有3种含义：________、________和________。

2. 统计总体具有下列3个特征，即________、________和________。

3. 在统计的研究方法中，________法、________法和________法是最基本的方法。

4. 统计上把________叫做变量。

5. 标志是反映________的特征的，可分为________和________；而指标是反映________的特征的，可分为________和________。

6. “2013年元旦小长假四川省接待游客总数达676.73万人次”是一项完整的统计指标，它主要由4部分构成，其中“2013年元旦小长假四川省”是________、“接待游客总数”是________、“676.73万”是________、“人次”是________。

7. 统计指标体系是指________，它可以分为两大类，即________指标体系和________指标体系。

8. 研究“城市按照综合竞争力排名”，最恰当的计量尺度是________。

9. 样本所取自的总体又称为________，样本又称为________。

10. 变量根据其变量值的取值情况可分为________和________。

五、应用能力训练题

1. 列出表1-1中各总体的总体单位、数量标志(2个)、品质标志(2个)。

表 1－1　相关统计概念训练题

总　　体	总体单位	数量标志	品质标志
大学生			
公司全体员工			
7 月生产的汽车			
暑期销售的计算机			
上年发生的交通事故			
商业网点			
手机			
旅游景点			
“两会”期间代表们所提的议案			
图书馆的藏书			

2. 当研究表 1－2 中的问题时，请确定 4 种计量尺度中最恰当的计量尺度。

表 1－2　计量尺度训练题

研究的问题	计量尺度
上班采用的交通方式	
每周的学习时间	
学生乘车到学校的行驶路程	
昨晚 10：00 看的电视节目	
城市按照综合竞争力排名	
社会保障编号	
大学课本的价格	
校园气温	
对学校食堂的评价	
学院足球队队员的身高	
大学毕业生毕业论文质量等级	
城市进出口总额	

3. 在为某大型购物商场进行市场调查期间，在持有该公司信用卡的所有顾客中，随机挑选了 300 名顾客进行电话询问，搜集到了有关的信息。其中，有一个调查问题是这样的：在过去的 6 个月里，你是否购买过该商场的商品？购买金额为多少？被询问者中有 10％回答“是”，且这些人在过去的 6 个月里，平均购买金额为 1 782.67 元。根据这个数据，经过统计方法分析，得到在一定可靠程度下，该商场持有信用卡的顾客购买比例及购买金额的区间范围。



（1）请指出本案例中如下统计概念：总体、总体单位、样本、样本量、品质标志、数量标志、数量指标、质量指标、统计量、参数。

（2）本案例采用的统计方法属于描述统计还是推断统计？

4. 下面是社会经济生活中常见的统计指标。

出口贸易总额　旅游收入　经济发展速度　人口出生率　失业率　城镇居民人均可支配收入　恩格尔系数　固定资产投资　消费者价格指数

这些指标中，哪些是数量指标？哪些是质量指标？如何区分数量指标与质量指标？

5. 关于2011年江苏入境游客抽样调查情况的通报如下。

由国家旅游局统一部署、江苏省旅游局组织实施的2011年度入境游客花费抽样调查工作，于2011年3～7月在南京、苏州、无锡、扬州和南通五市的部分旅游饭店、旅行社及南京禄口国际机场、南通边防检查站进行。共回收入境游客抽样调查问卷1 704份，超过21.7%。其中，口岸调查521份，占回收问卷的30.6%；饭店调查1 183份，占69.4%。在回收的调查问卷中，外国人1 120份，占65.7%；香港同胞217份，占12.7 %；澳门同胞56份，占3.3%；台湾同胞311份，占18.3%。

2011年全国入境过夜旅游者抽样调查结果显示：入境过夜旅游者的花费总水平比2010年有所提高。江苏的具体情况是：人均天花费为213.82美元/人天，同比增长2.0%。其中：外国人为213.12美元/人天，同比增长3.5%；香港同胞为216.21美元/人天，同比下降4.0%；澳门同胞为227.56美元/人天，同比下降3.9%；台湾同胞为213.39美元/人天，同比增长2.1%。

2011年全国入境游客抽样调查显示：2011年入境过夜旅游者在中国境内停留时间与2010年相比基本持平。其中，外国人、香港同胞和澳门同胞停留时间有所缩短，台湾同胞停留时间与2010年基本持平。2011年入境过夜旅游者在江苏省南京、无锡、苏州、南通、扬州五市的平均逗留时间分别为：2.2天、1.7天、3.0天、2.1天、2.3天，与2010年同期相比分别为：缩短0.3天、缩短0.2天、缩短1.3天、延长0.9天、延长0.2天。

（资料来源：http：//www.lygtour.gov.cn/article/82.html.）

分析指出上述案例中的总体、样本、变量、数量指标、质量指标。

6. 若对校园中大学生关心的话题进行调查研究（如大学生上网状况调查、就业分配去向及初职预期收入调查等），你能说出统计工作中所涉及的统计总体、样本、标志、指标、指标体系等概念吗？



第2章 统计调查

理论目标

(1) 理解统计调查的意义与种类。
(2) 掌握统计调查基本方式、方法及运用条件。
(3) 掌握统计调查方案的内容。
(4) 了解统计调查技巧。

能力目标

(1) 能设计统计调查方案。
(2) 能选择恰当调查方式、方法进行统计调查。

做一做

国际劳工组织(International Labor Organization)2013年1月21日发表2013年全球就业趋势报告(Global Employment Trend)显示，2012年共有1.97亿人失业，其中年轻人失业状况更为严重。预计2013年全球失业人数将继续上升。2013年，全国普通高校毕业生规模达到699万人，比2012年增加19万人，高校毕业生就业形势更加复杂严峻。新形势下，大学生们具有怎样的就业观念和求职意向?应该采取哪些措施对他们进行帮助和有针对性的指导?近来这个问题一直困扰着刚来大学生就业咨询服务中心工作的小张，他想就此问题对大四学生进行一次统计调查，摸清情况后才能有的放矢。小张该怎样进行统计调查?请帮助他设计一个周密细致的统计调查方案。



想一想

小张要进行的统计调查工作与一般社会调查工作不一致，他想直接在大学生群体中收集大量第一手数据。要想高质量完成这个庞杂的任务，首先要设计周密的调查方案，这是调查团队行动的纲要。调查方案越是缜密，调查工作的实施就会越顺利，调查数据质量越高。那么，一个完整的统计调查方案应该包括哪些内容、应该怎样设计呢？在进行调查时怎样选择和运用各种调查方式方法？统计调查有哪些技巧呢？本章将介绍这些问题。

2.1　统计调查概述

2.1.1　统计调查的意义和特点

统计调查(statistical survey)是根据统计研究的目的和任务，采用科学的调查方法，有计划、有组织地向客观实际搜集各种原始资料或次级资料的工作过程。

原始资料是指向总体单位直接搜集的，来源于直接的调查和科学实验的第一手统计数据(primary data)。在社会经济领域中，通常用调查的方法搜集必要的统计数据。例如，为进行宏观管理决策和科学研究，必须掌握最新的人口、国内生产总值、第二产业和第三产业的基本情况及人民生活变化情况，国家统计局系统和国务院各部、委、局的统计系统就承担着这些调查任务。在市场经济体制下，企业为在激烈竞争中取胜，需要快速准确地掌握市场需求和居民消费等信息，越来越多的调查公司和网络中心等承担起专门组织调查的任务。“做一做”中对高等院校大学生的调查就是用直接调查的方法搜集原始资料；而在自然科学和工程实际的各个研究领域，我们通常通过科学实验的方法获得统计研究的数据。例如，在医学领域，可通过临床试验数据分析某种药物或治疗方案的疗效，在农业科研中可通过实验的方法选取最优品种和最佳的种植方式等。

次级资料是指已经经过某个部门或地区加工整理过的，来源于别人调查和科学实验的第二手统计数据(secondary data)。对于社会上大多数的研究工作者和实际工作者来说，亲自去做直接的调查往往不可能或没必要，这时，可以通过各种渠道获取别人调查或科学实验的间接数据。例如，从统计年鉴、各种报表及报纸杂志上搜集资料，从因特网上搜集到世界各国有关资料等。随着网络技术的普及，次级资料的获得将变得越来越方便。由于次级资料一般都是从原始资料过渡而来，所以本节中，将重点介绍搜集原始资料的统计调查基本方式、方法和实施调查的方案。

统计调查是原始资料的主要来源方式，与其他社会调查相比较，它搜集的资料主要是各种数字资料，具有数量性特点；统计调查要对调查对象中的足够多的个体，甚至全部个体进行调查，从而具有大量性的特点；统计调查的目的是反映调查总体的数量特征，因此具有总体性特点。

2.1.2 统计调查的作用和要求

统计调查介于统计设计与统计整理工作过程之间，它是统计整理和统计分析的基础环节和前提条件，在整个统计研究中占有十分重要的地位。

1. 统计调查的作用

统计调查的主要作用表现在以下几个方面。

(1) 统计调查是认识社会的基本方法。

(2) 统计调查为国家管理和现代企业管理提供基本统计资料。

(3) 统计调查的资料是进行经济预测的重要依据。

(4) 统计调查是实施经济评价的重要基础。

2. 统计调查的要求

统计调查必须保证质量，才能正确反映客观事物，作为预测未来、实施经济评价和经济管理的依据。因此，对统计调查的基本要求是准确性、及时性、完整性、系统性和经济性。

(1) 准确性。准确性是指统计调查取得的各项资料必须真实可靠，符合客观实际，按事物的本来面貌如实反映情况。统计数字的真实性是统计工作的生命。统计调查的准确性不仅是技术性问题，而且是涉及坚持统计制度、统计法规的原则问题。我国统计立法的核心就是保障统计资料的准确性、客观性。

(2) 及时性。及时性是指在统计调查规定的时间内，尽快提供规定的调查资料，完成规定的各项调查任务。统计资料是进行管理决策不可缺少的依据，而客观经济现象又是不断变化的，因而统计数据具有很强的时效性，如果统计资料不及时，就成了“雨后送伞”，失去了价值。统计资料的及时性也是一个全局性的问题。任何一个调查单位不按规定的时间提供资料，都会影响全面的统计工作。因此，提高统计调查的及时性要求各单位必须有全局观念，有团队协作精神，遵守统计制度和规律。

(3) 完整性。完整性是指调查单位不重复、不遗漏，所列调查项目的资料搜集齐全。只有齐全的统计资料，才能比较正确地反映所研究的社会经济现象的全貌。

(4) 系统性。系统性是指综合资料中的各项统计数据应该配套，要能从不同侧面、不同层次上对调查对象的整体进行研究，能够从事物的内部结构和外部联系上进行对比分析。

(5) 经济性。经济性是指在满足一定准确度的要求下，要以最小的调查费用取得所需的统计资料。通常对调查数据的准确度要求越高，则调查的费用就越大。而任何一项统计调查总有一定的费用约束，因此，如果一味地强调资料的准确度，无视经济性要求，会造成不必要的人力、财力和物力的浪费。

这5项要求是相互结合、相互依存的，在每一次统计调查实践中，要根据具体情况综合考虑。一般而言，要以准为基础，力求在准中求快，以尽可能小的成本取得完整、系统的资料。



2.1.3 统计调查的种类

社会经济现象错综复杂，而且又处在不断变动中，这就决定了统计调查方法的多样性。统计调查可以从不同的角度，按不同的标志进行分类。

1. 按被研究总体的范围分类

按被研究总体的范围不同，统计调查可分为全面调查和非全面调查。

全面调查是针对有限总体而言的，指对构成调查对象的所有总体单位全部进行调查登记。例如，2013年，为了了解我国第二、第三产业的经营情况而进行的第三次全国经济普查，就属于全面调查。

非全面调查又称部分调查，是在总体中选择一部分个体进行调查，从所了解的局部数据了解总体情况。例如，为了了解某地区居民的消费水平情况，可以只搜集各个收入阶层的一部分居民消费方面的实际资料；为了掌握进出口商品的质量，可以取出一部分产品做质量检验。与非全面调查密切相关的是如何选择部分个体的问题。从总体中，恰当地选取部分个体进行调查，是统计学中最重要的问题之一。非全面调查有抽样调查、重点调查和典型调查等调查方法。

在国外，非全面调查都称为抽样调查(sampling)，分为概率抽样(probability sample)和非概率抽样(non-probability sample)。我们通常所说的抽样调查是概率抽样。

2. 按调查登记时间的连续性分类

按调查登记时间的连续性不同，统计调查可分为经常性调查和一次性调查。

经常性调查是随着被研究对象的变化，连续不断地进行调查登记。例如，企业的产品产量、公司的利润额、商品的销售量等，必须在观察期内连续登记。可见，连续调查的资料是说明现象的发展过程，体现现象在一定时期的总量。

一次性调查是指间隔一定时间的不连续调查。例如，商业网点数、高等学校数、商品库存量等现象，短时期内不发生什么变化，一般隔一段时间进行一次调查。因此，一次性调查并不意味着只能对现象调查一次，只是没有必要进行经常性的调查，只需间隔一定时间了解现象在某一定时点上的状况。

3. 按搜集资料的方法分类

按搜集资料的方法不同，统计调查可分为直接调查和间接调查。

直接调查(survey at first hand)是指直接搜集第一手统计资料的统计调查，如问卷调查、专家访谈、电话调查等；间接调查(survey at second hand)是指间接地搜集第二手统计资料的调查，如搜集报纸、杂志、电台、调查报告等现成统计资料。

以上各种分类是交叉的关系，而不是相互排斥的关系。例如，人口普查既是全面调查，又是一次性调查；进出口商品质量检验既是直接调查，又是非全面调查。实际工作中的每一项具体调查都可以从各个角度去认识。

2.2 统计调查方案

统计调查是一项涉及面广、程序步骤多、要求严格的科学工作。一项全国性的统计调查，往往需要动员成千上万的人协调工作才能完成。为了顺利完成调查任务，在调查之前需要设计一个调查方案，使统计调查工作有计划、有组织地进行。一个完整的统计调查方案应该包括以下基本内容。

1. 调查目的

调查目的就是通过统计调查所要达到的具体目标。只有确定调查目的，才能解决为什么要调查的问题，才能确定搜集资料的范围和方法，才能有效地组织统计调查工作。调查目的的确定应该具体明确、突出中心，否则，可能产生以下结果：匆忙辛苦地进行了调查，取得的资料可能是不需要的；或者材料的口径范围不一致，无法进行汇总得到综合结论，而真正需要的材料，反而没有搜集。这样不仅浪费人力、物力，而且也将延误工作。例如，在“做一做”中，小张的调查目的是要了解本校大四学生就业观念和求职意向，为制定措施进行管理工作提供依据。

2. 调查对象和调查单位

有了明确的调查目的，就可以确定调查对象和调查单位(survey unit)。调查对象是根据调查目的而确定的被调查的统计总体，统计总体这一概念在统计调查阶段称调查对象，不能把调查对象理解为被调查的个体；调查单位是构成调查对象的每一个总体单位，总体单位这一概念在统计调查阶段称调查单位，调查单位是进行调查登记的标志承担者。不要把调查单位理解为从事调查工作的工作部门或单位。

在调查方案中，有时还要规定统计资料的填报单位。填报单位也叫报告单位，它是填写调查内容、提交调查资料的单位。填报单位与调查单位有可能一致，也有可能不一致。例如，工业企业进行产品质量检验时，调查单位是工业企业的单个产品，报告单位是工业企业，这时，调查单位与填报单位是不一致的；而在工业普查中，调查单位是每一个工业企业，报告单位也是每一个工业企业，两者完全一致。在“做一做”中，调查对象就是全体大四学生，每一位大四学生是调查单位，同时也将是填报单位。

3. 调查项目和调查表

调查项目是指对调查单位所要调查的内容，是调查单位所承担的基本标志。例如，在人口普查中的调查项目有姓名、性别、年龄、民族、文化程度、职业等。在确定调查项目时应注意以下几点。

(1) 调查项目的内容要少而精，所选项目满足调查目的需要并且能确实取得资料。

(2) 调查项目的含义要明确、具体，切忌模棱两可。

(3) 调查项目的设置要考虑彼此间相互联系及同类调查的纵向衔接。

调查项目可以用调查问卷的形式表示，也可以设计成各种调查表。



调查问卷按传递方式不同，可分为报刊问卷、邮寄问卷、送发问卷、访问问卷和网上问卷；按问卷填答者的不同和调查方法的不同，可分为自填式问卷和代填式问卷。不同的调查问卷在具体结构、题型、措辞、版式等方面会有所不同，但在结构上一般由开头问候语部分、中间主体部分和背景部分及最后的结束语部分组成。

调查问卷中问题的设计有两种，即开放式问题和封闭式问题。开放式问题在提出问题时不提供任何答案，由被调查者自由填写。这种调查问卷，主要用于希望了解的内容比较详细的情况。封闭式问题在提出问题的同时，还必须将答案设计出来，主要用于了解被调查者的基本意向。这是现代问卷调查中采用的主要形式。

调查问卷的设计可以按以下步骤进行：首先，明确调查目的，把握调查主题，确定调查的项目；其次，针对每一个调查项目，设计若干问题，确定问句类型，进行初步设计；再次，对初步设计的问卷在小范围内进行实验性调查，发现问题，及时修改，使问卷完善；最后，确定须印制的份数，交付印刷，制成正式调查问卷。

调查表(survey questionnaire)是调查项目的表现形式，其作用在于能够条理清晰地表述调查内容，便于登记调查资料。

调查表由表头、表体和表脚3部分组成。表头在调查表的上方，标明调查表的名称、填报单位的名称等。表体是调查表的主体部分，由表格、调查项目等组成。表脚在调查表的下方，包括调查人员或填报人员签名、审核人员签名、填报日期等。

调查表的形式有单一表和一览表两种。单一表是在一张调查表上只登记一个调查单位的资料，可以容纳较多的调查内容，且便于分类和整理。一览表是在一张调查表上可以登记很多的调查单位的资料，却不能容纳较多的调查内容。

在“做一做”中，可以将调查项目设计成简单的调查问卷如下：

××学院大四学生就业意向调查问卷

亲爱的同学：

您好！

开展此调查旨在了解我院大学生就业观念、求职意向等问题，为更好地满足同学们择业的需求提供依据。您的建议和答卷对我们很重要，本调查采用无记名形式，不会给您个人带来任何不利影响，请实事求是回答。我们对您的合作表示衷心的感谢！

××学院学工办调查小组
×年×月

1. 您最理想的工作岗位是在：

A. 政府部门　　B. 企事业单位　　C. 外资企业

D. 民营企业　　E. 自己创业

2. 您择业最看重的是：

A. 薪酬　　B. 专业对口　　C. 实现自我价值　　D. 兴趣、爱好

E. 就业地区　　F. 工作环境

3. 就您目前所学的专业知识，您有信心找到工作吗？

A. 很有信心　　B. 有信心

C. 不是很有信心　　D. 一点信心也没有

4. 您认为比较理想的就业地区是：
A. ××市区　　B. ××周边区市　　C. 省内其他地区
D. 外省市　　E. 由父母及家人决定
5. 您期望的月薪是________，对月薪最低要求是________。
A. 2 000 元以下　　B. 2 000～3 000 元　　C. 3 000～4 000 元
D. 4 000～5 000 元　　E. 5 000 元以上
6. 你认为招聘方最看重：
A. 学校名声　　B. 专业优势　　C. 个人综合素质　　D. 学历
7. 影响就业的因素有很多，那您认为最关键的因素是：
A. 个人能力　　B. 社会背景　　C. 机遇
D. 学校品牌　　E. 其他
8. 您获取就业帮助的渠道是：(多选)
A. 网络　　B. 新闻媒体　　C. 学校　　D. 招聘会
E. 老师　　F. 亲友帮助　　G. 其他
您的性别：A. 男　　B. 女　　您的专业：
对您的合作再次表示真诚的感谢!
1. 您填答完这份问卷后有何感想? ____________________
2. 您对问卷有什么建议和要求? ____________________

4. 调查时间和调查期限

调查时间是指调查资料所属的时间。如果所要调查的是时期现象，调查时间就是资料所反映的起讫日期；如果调查的是时点现象，调查时间就是规定的统一标准时间。统计调查中，应明确规定调查时间，以满足统计调查的准确性要求。

调查期限是指进行调查工作的起止时间，包括搜集资料和报送资料整个工作所需的时间。统计调查中，应根据任务量的大小及人力、物力、财力等情况尽可能缩短调查期限，以满足统计调查的及时性要求。

例如，对某市 2012 年批发和零售业商品销售情况进行调查，即对商品销售额、批发总额、零售总额等指标进行调查，这个调查时间是时期，即 2012 年这一年。从 2013 年 1 月 1 日起开始调查，截至 2013 年 1 月 31 日将资料搜集、整理完毕，则调查期限为 1 个月。又如，某管理局要求所属企业在 2013 年 1 月 15 日上报 2012 年年末从业人员资料，则调查时间是标准时点 2012 年 12 月 31 日，调查期限是 15 天。

5. 调查的组织和实施

为保证调查工作的顺利进行，在调查方案中还应该有一个周密的组织实施计划。其主要内容包括调查工作的领导机构和办事机构、调查人员的组织和业务培训、调查经费的预算和开支方法、工作进度安排等。

例如，2013 年我国开展第三次全国经济普查，普查方案主要包括以下内容。

一、普查的主要目的

全面调查了解我国第二产业和第三产业的发展规模及布局，了解我国产业组织、产业结构、产业技术的现状及各生产要素的构成，进一步查实服务业、战略性新兴产业和小微企业的发展状况，摸清我国



各类单位的基本情况，全面更新覆盖国民经济各行业的基本单位名录库、基础信息数据库和统计电子地理信息系统中国城市低碳经济网通过普查，进一步夯实统计基础，健全统计工作的部门协调机制和信息共享机制，为加强和改善宏观调控，加快经济结构战略性调整，科学制定中长期发展规划，提供科学准确的统计信息支持。

二、普查的对象和范围

第三次全国经济普查的对象是在我国境内从事第二产业和第三产业的全部法人单位、产业活动单位和个体经营户。具体范围包括采矿业，制造业，电力、热力、燃气及水生产和供应业，建筑业，批发和零售业，交通运输、仓储和邮政业，住宿和餐饮业，信息传输、软件和信息技术服务业，金融业，房地产业，租赁和商务服务业，科学研究和技术服务业，水利、环境和公共设施管理业，居民服务、修理和其他服务业，教育、卫生和社会工作，文化、体育和娱乐业，以及公共管理、社会保障和社会组织等。

三、普查的内容和时间

普查的主要内容包括单位基本属性、从业人员、财务状况、生产经营情况、生产能力、原材料和能源及主要资源消耗、科技活动情况等。

普查标准时点为2013年12月31日，普查时期资料为2013年年度资料。

四、普查的组织和实施

第三次全国经济普查是一项重大的国情国力调查，各地区、各部门要按照“全国统一领导、部门分工协作、地方分级负责、各方共同参与”的原则，突出重点，优化方式，统一组织，创新手段，认真做好普查的宣传动员和组织实施工作。

为了加强对普查工作的组织和领导，国务院将成立第三次全国经济普查领导小组，负责普查组织和实施中重大问题的研究和决策。普查领导小组由国务院领导同志任组长，成员单位包括国务院办公厅、统计局、发展改革委、中央宣传部、中央编办、监察部、民政部、财政部、税务总局、工商总局和质检总局等部门(组成人员名单另发)。普查领导小组办公室设在统计局，负责普查的具体组织实施和协调。其中，涉及普查经费方面的事项，由财政部负责和协调；涉及固定资产投资保障方面的事项，由发展改革委负责和协调；涉及企业和个体工商户名录方面的事项，由工商总局、税务总局负责和协调；涉及机关和事业单位名录方面的事项，由中央编办负责和协调；涉及社团、基金会、民办非企业单位及基层自治组织名录方面的事项，由民政部负责和协调；涉及组织机构代码方面的事项，由质检总局负责和协调；涉及各级政府及其普查工作人员在普查工作中违法违纪行为的事项，由监察部负责和协调处理。国务院其他各有关部门，也要按照各自的职能，各负其责、通力协作、密切配合。

地方各级人民政府要设立相应的普查领导小组及其办公室，加强领导，认真组织好本地区的普查实施工作。对于普查工作中遇到的困难和问题，要及时采取措施，切实予以解决。要充分发挥街道办事处和居民委员会、乡镇政府和村民委员会的作用，广泛动员和组织社会力量积极参与并认真配合做好普查工作。地方普查机构应当根据工作需要，聘用或者从有关单位商调符合条件的普查指导员和普查员，并及时支付聘用人员的劳动报酬，保证商调人员在原单位的工资、福利及其他待遇不变，稳定经济普查工作队伍，确保普查工作顺利进行。

五、普查的经费保障

第三次全国经济普查所需经费，由中央和地方各级人民政府共同负担，并列入相应年度的财政预算，按时拨付、确保到位。

六、普查的工作要求

坚持依法普查。所有普查对象必须严格按照《中华人民共和国统计法》和《全国经济普查条例》的规定，按时、如实地填报普查表。任何单位和个人不得虚报、瞒报、拒报、迟报，不得伪造、篡改普查数据。地方各级人民政府统计机构和监察机关要加大对普查工作中违法违纪行为的查处力度，坚决杜绝人为干扰普查工作的现象，确保普查工作顺利进行和普查数据质量。普查取得的单位和个人资料，严格

限定用于普查目的，不作为任何单位对普查对象实施处罚的依据。各级普查机构及其工作人员，对在普查中所知悉的国家秘密和普查对象的商业秘密，必须履行保密义务。

充分运用现代信息技术。利用统计电子地理信息系统，全面建立普查区电子地图；巩固和拓展统计联网直报系统成果；积极推广使用手持电子数据采集设备，努力提高普查工作的信息化水平和效率，减轻基层普查人员的工作负担。

加强宣传工作。各级普查机构应会同宣传部门认真做好普查宣传的策划和组织工作，主动向新闻单位提供情况。报刊、广播、电视和互联网等媒体要广泛深入宣传经济普查的重要意义和要求，宣传普查工作中涌现出的典型事迹，报道违法违纪案件查处情况，引导广大普查对象依法配合普查，教育广大普查人员依法开展普查，为普查工作顺利实施创造良好的舆论环境。

（资料来源：http：//finance.chinanews.com/cj/2012/11-12/4322360.shtml.）

2.3 统计调查方式

在明确了统计调查目的、调查内容等之后，就涉及采用什么方式实施调查的问题。统计调查的方式很多，本节介绍几种常用的重要统计调查方式。

2.3.1 普查

1. 普查的意义和特点

普查(census)，即普遍调查，是依据调查任务而专门组织的一次性的全面调查。它可以取得比较准确的全面统计资料。通过普查，从宏观上看，可以摸清一个国家的国情和国力，为了解一个国家的人力资源和物质资源的现状及利用情况，制定重要政策和长期发展规划提供重要依据；从微观上看，普查也可用于某些小范围的市场调查，如对市场上某种商品的库存、供应和销售的全面调查，为掌控商品行情、进行经营管理提供重要决策参考。普查具有以下 3 个主要特点。

(1) 一次性。普查一般用来调查属于一定时点上现象的总量，这些时点现象的数量在短期内往往变动不大，不必做连续登记，只需间隔一段较长时间进行一次性的调查。而且，普查的规模大、指标多、任务重，耗费大量人力、物力和时间，不可能采用经常性调查，只能采用一次性调查。例如，人口普查工作，不可能年年进行，更不可能月月进行，我国 2010 年进行的第六次人口普查与 2000 年的第五次人口普查就间隔了 10 年。

(2) 时点性。普查的对象主要是时点现象，所以每次普查都有标准时点。例如，2013 年第三次全国经济普查的标准时点为 2013 年 12 月 31 日。但是值得指出的是，普查也不排斥对某些时期现象在某段时期内数量表现的调查。例如，第三次全国经济普查还包括对零售批发业在 2013 年度的销售总额的调查等。

(3) 全面性。普查对象范围广，调查内容详细，所以它比其他任何方式的调查更能掌握全面、详尽的统计资料，具有重要的分析价值。例如，我国第三次全国经济普查有不同种类的普查表，涉及的调查指标众多。内容包括了单位基本属性、从业人员、财务状况、生产经营情况、生产能力、原材料和能源及主要资源消耗、科技活动情况等，为国家制定国民经济和社会发展规划，为各级部门提高决策和管理水平提供了重要信息。



2. 普查的组织实施

普查的组织形式有两种。一种是自上而下组织专门普查机构，配备一定数量的普查人员，对调查单位直接进行调查登记，如我国的人口普查都采用这种形式。另一种是自下而上由被调查单位填报调查表格逐级上报实施普查，如我国的物资库存普查。但即使是后一种形式，也仍需组织普查的领导机构，配备专门人员对普查工作进行组织领导。根据普查特点，在组织实施时应按系统工程要求，严把质量关，具体实施包括以下过程。

(1) 建立普查的组织领导机构。普查范围广，调查单位多，花费时间长，要求必须建立各层次组织机构，安排专门人员参加普查工作，并且职责明晰、分工明确。

(2) 确定统一的标准时间。这是指规定某一时刻或现象所属的某一段起止时间作为登记普查对象的标准时点或标准时期。这样才能保证所有调查单位的调查资料都是反映同一时间特征的，避免所搜集的资料因时间不同，汇总后不能反映客观事实而失去准确性。例如，我国第六次全国人口普查规定的标准时点为2010年11月1日零时，而相差一天，我国的人口总量就会有几万人的差异。当今社会的高速变动性使得在普查时确定统一标准时间、严格保证调查资料从属这一时间至关重要。

(3) 规定各阶段工作进度。例如，我国在2010年进行的第二次全国R&D资源清查工作分为6个阶段。2009年3～9月为前期准备阶段，2009年4～11月为方案研制阶段，2009年6～12月为布置培训阶段，2010年1～9月为填报汇总阶段，2010年9～12月为资料开发阶段，2010年12月～2011年2月为总结表彰阶段。有了工作进度，才能保证普查按计划同时开始、按期完成，使调查资料得以及时汇总，发挥统计调查的时效性价值。

(4) 规定普查的项目和指标。普查的项目必须统一，其内容、解释口径、计算方法等一经确定，不得任意增减改变，以免影响汇总综合，降低资料质量。同一种普查，历次项目和指标应尽可能保持连贯性，以便更好地进行历次调查资料的对比分析，观察某种现象变化发展的情况，认识客观发展规律。

(5) 制定严格质量控制方法。普查工作是一项庞大的系统工程，普查资料的质量控制是贯穿始终的重要工作。它包括普查前的质量控制、普查中的质量控制和普查后的质量控制，只有步步为营实施控制，才能保证普查资料的准确性，提高普查工作的质量。

3. 我国的周期性普查

普查作为获取有关国情、国力资料的重要统计方法，在我国已经纳入了规范化、法制化的轨道，我国政府也制定了“以周期性普查为基础”的整体统计调查方法体系。我国的周期性普查制度确立于1994年。已有的重大普查工作的统筹安排为人口普查，逢0的年份实施，每10年1次。人口普查的对象是随着社会经济的发展有所变化的。其中2010年进行的第六次人口普查的对象是在中华人民共和国境内居住的自然人，以及在中华人民共和国境外但未定居的中国公民；全国经济普查每10年进行两次，分别在逢3、逢8的年份实施。经济普查以从事第二、第三产业的全部法人单位、产业活动单位和个体经营户为对象；农业普查，逢6的年份实施，每10年1次。农业普查以从事第一产业活动的单位和农户为对象。

2.3.2 概率抽样调查

1. 概率抽样调查的意义和特点

概率抽样调查(sampling survey)又称随机抽样调查，简称概率抽样，是一种非全面调查。它是按照随机原则从总体中抽出适量单位组成样本予以调查，用样本资料推算总体数量特征的一种调查方法。

例如，从一批产品中随机抽取100件进行质量检验，计算出这100件产品合格率，然后据此推断出这一批产品的合格率。

抽样调查是目前国际公认、世界各国普遍采用的一种科学有效的统计调查方法。它有以下3个主要特点。

(1) 它是非全面的“代表性”调查。概率抽样调查通过样本的调查推断总体的特征，这种推断虽存在一定的抽样误差，但误差的范围是可以计算和加以控制的，使推断结果具有一定可信程度。这一特点使它不同于全面调查，也与其他非全面调查有显著区别。

(2) 概率抽样调查按照随机原则取样。所谓随机性原则，就是总体中调查单位的确定完全由随机因素决定，单位中选与否不受主观因素的影响，保证总体中每一单位都有同等的中选可能性。采用随机原理抽选样本调查单位，能保持样本的结构近似于总体的结构，使样本对于总体具有代表性。

(3) 概率抽样调查每个个体被抽中的概率是已知的，或者是可以计算出来的。这有利于科学地把握调查的准确性。

概率抽样调查法在现代统计调查中有准确性高、时效性强、经济性好、应用性广的显著优越性。它已成为统计调查方法的主体，本书将在后面的第6章加以详细讨论。

2. 我国的抽样调查实践

我国在20世纪50年代，在城市职工住户调查、农村经济调查中曾经采用过抽样调查，后中断。改革开放后，在借鉴国外抽样方法的基础上，结合我国实际情况，国家统计局在80年代组建城市经济调查总队、农村经济调查总队，在1998年成立企业调查总队，各地方也相应有城调队、农调队和企调队，他们为国家成功地进行了工业产品质量抽样检验、城市住户调查、市场抽样调查、人口抽样调查、农产量抽样调查，等等。

随着经济的迅猛发展，快速、高效地获取准确的统计资料已成为企业在激烈竞争的社会中生存的重要手段。抽样调查在社会经济各部门得到了非常广泛的使用，许多人通过参与调查或被调查感受到“统计就在我们身边”。

2.3.3 重点调查

1. 重点调查的意义

重点调查是一种非全面调查，属于非概率抽样。它是在全部调查单位中选择一部分重点单位进行调查，借以了解总体的基本情况的一种调查。所选择的重点单位是指这样一些



单位，它们的数目在全部单位数中只占很小的比例，但它们的标志值的总和在总体的标志总量中却占很大的比例，通过对这部分单位进行调查，就能从数量上大致反映总体的基本情况。

例如，要了解我国出口商品总额的基本情况，以美国、欧盟、中国香港地区、日本、东盟、韩国及中国台湾省几个大的贸易伙伴为重点进行调查即可完成。虽然它们在我国所有的贸易伙伴中是少数，但它们是大的贸易伙伴，贸易额占绝对大的比例。对这些重点贸易伙伴进行调查，比全面调查要省时省力，而且更能及时地了解全国出口贸易总额的基本情况。

2. 重点调查的特点和应用

(1) 重点单位的选择不带有主观因素。它只着眼于所研究现象主要标志的总量，而与平常所说“重点单位”有区别。只要调查单位的主要标志总量不占绝大比例，仅因为技术创新、管理先进或特殊原因而被列为重点管理的，都不列为重点调查单位的范畴。

(2) 重点调查不能精确推断总体总量。它只对总体进行概括性的推断。

重点调查能以较少的投入、较快的速度高效率获取某些现象主要标志的基本情况或变动趋势。它的适用范围很广。当调查目的只要求了解总体的基本状况和发展趋势，不要求掌握全面数据，同时调查少数重点单位就能满足需要时，采用重点调查就非常适宜。目前，国家统计局每月公布的70个大中城市住宅销售价格数据，就是国家为了及时了解全国城市商品房价格的变动趋势，对全国70个大中城市的商品房价格的变化进行调查得到的，这种调查就是重点调查。我国有一些重点调查已列入定期报表制度，以便及时取得必要的资料，如国家统计局近年来积极推进的企业一套表联网直报制度。

根据管理学界所熟知的帕累托80/20法则，当今社会，在企业经营管理中和社会生活中普遍存在着不平衡的现象，这为我们发挥重点调查作用，了解总体概况提供了广阔的实践领域。

2.3.4 典型调查

1. 典型调查的意义

典型调查(representative sampling)也是一种非全面调查，属于非概率抽样。它是根据调查的目的任务，在对所研究的现象总体进行初步分析的基础上，有意识地选取有代表性的典型单位进行的调查。

例如，为了研究大中型国有企业改革成功的经验和失败的教训，从全部大中型国有企业中选出若干先进的典型和失败的典型，分别进行调查比较，在此基础上，认识大中型国有企业改革的规律和本质。

典型调查能否取得良好效果，关键在于能否正确选择典型单位。所谓典型单位，是指在同一事物中能最充分、最突出地体现总体共性的代表单位。选择典型单位时应注意，要在对调查对象总体进行全面分析，对可供选择单位进行反复对比基础上选择代表性较强的单位。



2. 典型调查的特点和应用

(1) 典型单位的确定是人为选择的，取决于统计调查人员对调查对象全面分析基础上主观的判断选择。这与重点调查中重点单位的选取是不相同的。

(2) 一般说来，典型调查的目的是发现问题、观察趋势，而不能不加限制地直接以调查结果推断总体。典型调查经常用来验证全面调查(如普查)数据的质量，而且典型调查的深入了解也可以作为全面调查的补充说明。

应用典型调查时，选取典型单位的方法可以灵活多样。例如，为了研究获得成功的经验或总结失败的教训，应该选择调查对象中先进或落后的作为典型单位；为了了解总体的一般数量表现或一般发展趋势，可选择中等的带有普遍性的那部分单位作为典型单位。若调查总体各单位发展条件较一致，彼此之间差异小，可选一个或几个典型单位进行调查，即通常所说“解剖麻雀”式典型调查；若各单位之间发展条件差异大，涉及问题较复杂，则应按一定标志，将总体划分为几个类型，从各类中按比例选取典型单位调查，即通常所说“划类选典”式典型调查。

典型调查不是统计工作过程特有的方法，但它在统计活动中有重要的适用领域。一般说来，当我们调查的目的主要不在于取得现象的总体数值，而在于了解与统计数字有关的生动的具体情况，即与现象有关的各种社会条件及其之间的相互联系，以便进行深入的定性与定量相结合分析时，经常运用典型调查的方法。应该注意的是，由于典型单位是“有意识地选取有代表性的单位”，因而人们的主观认识会产生较大影响。如果调查者研究问题的态度科学客观，对被研究对象情况熟悉，运用典型调查能够取得满意效果；反之可能会产生较大偏差。因此，在社会经济实践中，典型调查经常与其他调查结合起来使用，避免出现结论的片面性。

2.3.5 统计报表

1. 统计报表的意义和资料来源

统计报表(statistical report form)是按国家统一规定的调查文件，以基层统计的原始记录为依据，自下而上、逐级定期地提供统计资料的一种调查方式。它是一张调查表，报表中的指标项目就是调查项目。

统计报表的资料来源于基层单位的原始记录。原始记录是通过一定表格形式对基层生产经营活动所做的最初记载，是反映社会经济活动的第一手材料。原始记录种类繁多、范围广泛。例如，企业产品产量记录、原材料入库记录、职工出勤和工作记录、商品销售记录等，都是原始记录。

从原始记录到统计报表，中间还要经过统计台账和企业内部报表的填制环节。统计台账是基层单位根据核算和填制统计报表的需要，按时间顺序登记原始记录的一种账册。它是为积累和整理资料而设置的，是从原始记录到统计报表的中间环节。例如，企业每天会有许多的产品入库单，这些产品入库单就是企业的原始记录；每天把这些入库单汇总得到该产品的当日产量，按日期登记在一本账册中，即统计台账；将统计台账中每日产量累



计，得到当月产品产量，据此填入统计报表中的产品产量。可见，原始记录、统计台账和统计报表之间联系密切、层次递进。

2. 统计报表的特点和应用

统计报表曾经在我国高度集中的计划经济年代是一种最主要的统计调查方式。我国在多年的统计实践中，最大限度地发挥了统计报表的作用，成为世界上运用统计报表最成熟的国家之一。在当前的市场经济条件下，虽然由于统计报表费时费力，中间环节多，不再作为获得统计资料的主要和唯一的形式，但在某些方面统计报表仍有不可替代的作用。

(1) 统计报表在普查中应用。普查是全面调查，其调查对象是组成社会最基本的单位。这种从基本单位到整体的调查、搜集资料并逐级上报汇总的过程，就离不开统计报表制度，只有通过报表制度的规范性，才能使普查工作严格按程序进行，保证普查资料的质量。

(2) 统计报表在其他调查中应用。在抽样调查和重点调查中，为规范调查资料的填写上报程序，也需要建立统计报表制度，统计报表在非全面调查中同样得到广泛应用。

随着我国统计信息技术现代化体系的建立，我国的统计报表在报送程序和报送手段上已发生了深刻的变革。传统的手工报表正被计算机的信息处理所替代，传统的邮寄电信报表正被远程传输及信息网络所替代，传统的自下而上、逐级上报的程序随着网络的建立也可转变为“直达车”形式。

3. 企业一套表

企业一套表是一种以企业为基本统计单元，将需要采集的统计调查内容集中设置并统一布置和收集的统计报表制度。它主要根据调查对象所从事的国民经济行业活动进行设计，内容包括调查对象的基本情况、主营业务及附属活动。现行企业一套表制度规定：统计报表填报单位为法人单位。

企业一套表按照统一设计、统一标准、统一调查单位、统一布置的原则实施，对已经相对落后并在一定程度上制约统计科学发展的传统生产方式、管理体制和工作机制进行了重大变革。即由分散设计转变为统一设计，由分散布置转变为统一布置，调查单位确定由各专业自行确定转变为统一确定，原始数据采集由国家间接采集、各级统计机构层层上报转变为国家直接采集、各级统计机构在线同步共享。

2009 年，国家统计局启动了企业一套表的试点工作。2011 年，国家统计局制定并下发了《全面实施企业(单位)一套表工作总体规划》。根据这个规划，企业一套表实施工作分 3 个阶段。第一阶段，从 2011—2012 年，在规模以上工业企业、资质内建筑业企业、限额以上批零住餐企业等“三上”企业，以及房地产企业、年综合能源消费量 1 万吨标准煤及以上的建筑业企业和三产单位首先实施。第二阶段，在 2013 年进行实施。将一套表的实施范围扩大到工业、建筑业、批发和零售业、住宿和餐饮业等国民经济行业的抽样调查企业，并将劳动工资等专业统计纳入一套表范围。第三阶段，从 2014—2015 年，要全面实施一套表，要将生产价格、服务业统计等抽样或重点调查内容纳入一套表范围，初步建立起国家核心统计指标体系和基于元数据标准的统计调查报表管理体系。

以上是从不同角度对统计调查方式的分类，各种统计调查方式均有其不同的特点、局限性，因而各有其适用范围。在实际工作中，应按照建立以周期性普查为基础，以经常性抽样调查为主体，以必要的统计报表、重点调查、典型调查、综合分析等为补充的统计调查方法体系的要求，针对不同的调查对象、调查目的和任务，灵活运用多种调查方式，才能准确、及时、全面、系统地提供统计资料。

2.4 统计调查的方法和技巧

统计调查方法是指统计工作确定调查方式之后，向客观实际搜集资料的具体方法。本节从传统调查方法和网上调查两个方面加以介绍。

2.4.1 传统调查方法

1. 观察法

观察法又称直接观察法，是由调查人员亲自到现场对调查对象进行观察、记录和计量，以取得第一手统计资料的调查方法。例如，对超市购买者进行调查时，为获得出向和入向人数，则调查员亲自到超市的出入口实际观察，并进行记录。这种方法能够保证统计资料的相对准确性。但是这种方法需要大量的人力、物力和时间，在任务紧迫情况下，不宜采用。而且对历史资料进行调查时，不可能通过直接观察搜集资料。另外，这种方法容易受到观察者和被观察者主观因素的影响而产生各种观察误差。例如，当人们知道在被观察时会改变行为，观察者在现场观察一定时间后会因疲倦产生观察误差，这些都会降低调查资料的质量。

2. 访问法

访问法又称采访法、询问法，是由调查人员通过口头、书面等方式向被调查者了解情况，取得第一手资料的调查方法。例如，上面提到的在对超市购买者进行流动人数调查的同时，对消费者的购物金额进行调查，调查人员在出入口对顾客进行询问，根据回答一一记录，并填写相应的调查卡片或表格据以搜集统计资料的调查方法。

常用的访问法有以下 3 种。

(1) 面谈访问法。由调查者直接与被调查者接触，通过有目的的当面交谈获得资料的一种方法。面谈访问调查简便灵活，可随机应变提出问题且调查表回收率高，但调查成本高，调查范围有限，且调查结果易受调查人员综合素质的影响。

面谈访问法可进一步分为入户访问和街上拦截访问两种情况。

入户访问法，需要访问员直接深入调查对象的居住地或工作场所与被调查者接触，利用访问式问卷逐个问题询问并记录下对方的回答；或者是向被调查者阐明填写方法后，将问卷交给被调查者由对方自填。

街上拦截访问法，是指访问员在街头随机拦截调查对象，就所调查的问题进行面谈访问的方法。可以采用纸质访谈和电脑访谈两种形式。后者的做法可以是利用手提电脑当街



填写电子问卷，也可以是在街头拦截调查对象后，请到附近访问室，让其坐在电脑前，访问员启动电脑里的调查问卷程序，对被调查者进行简单的操作说明，被调查者回答屏幕上问卷。

（2）电话访问法。这是一种通过电话向被调查者询问调查内容的方法。电话访问能迅速获得有关资料，成本低，特别是在那些不容易见面访问或被调查者不愿意接受访问情况下，采用电话访问，有可能获得调查成功。但是，由于电话访问对象只限于通话者，调查范围受到极大限制。且通话时间不宜过长，图表、照片无法显示，故一般适用于调查项目单一、问题简单明确、要求快速获得所需调查信息的情况。

电话访问法通常可以采用电话号码簿抽样法和随机拨号法抽取概率样本，来确定被调查者。

电话号码簿抽样法，首先找出与被研究总体相对应的最新的电话号码簿，将电话号码簿上的电话号码建为一个计算机文件，采用随机数表随机抽出所需的单位数。也可以先用随机数表从电话号码簿中抽出若干页，然后分别在每一页中再用随机数表随机抽出若干电话号码。

随机拨号法，首先列出被研究总体所在的所有地区的电话局号，在每一个电话局号下利用随机数表产生若干组4位随机数，电话局号加上4位随机数即构成随机抽出的电话号码。拨打电话进行调查，若该号码为空号或者是工作单位号码(假设进行的是居民调查)，则在该电话号码末尾4位随机数相邻的上下各两个(或者3个、4个……)4位数字中，另外再随机抽取一个4位数字，再与前面的局号连成一个新电话号码拨打。随机拨号法不需要电话号码簿，但易生成许多无关号码而浪费调查时间。

电话访问法可以采用中心控制电话访问和电脑辅助电话访问两种方法。

中心控制电话访问，通常通过一套带有监听系统的专门设备进行。访问员拨通电话后，按照问卷逐一提出问题，同时用铅笔将对方答案随时记录下来。督导员利用监听装置进行监听(一名督导员可同时监听10～20名访谈员)，掌握访谈过程实况，并及时予以纠正。

电脑辅助电话访问，通常访谈员利用电脑读出屏幕上问题进行提问，并同时输入对方相应的答案，计算机会自动显示恰当的下一道问题。这种方法在提问时，电脑可以自动跳过应当略去的问题，并且可以随时进行数据整理和数据分析工作。

（3）邮寄访问法。将设计好的问卷通过邮寄方式送到被调查者手中，说明填表要求和方法，由被调查者填写后寄回，以获取信息资料的方法。这种方法调查面广、成本低、访问对象填写自由方便，但调查表的回收率低、回收时间长，调查结果的完整性往往不易控制。

邮寄访问法的问卷可以是纸质问卷，也可以是磁盘。

采用邮寄访问法进行调查发出的邮寄问卷一般包括信封、封面信、问卷、回邮信封(贴邮票)、礼品等。被调查者完成问卷后用回邮信封寄回。

3．实验法

实验法指调查人员根据研究目的，通过实验对比，对调查对象的某些因素之间的因果

关系及其发展变化过程，进行实验观察和分析，以取得调查资料的方法。例如，若要了解饮料配方的改变对销售量的影响情况，则需选定一个地区范围，将新旧两种配方的饮料投入市场进行试验对比，观察其销售量变化和消费者反映，获得数据作为是否采用新配方的依据。

实验法能够排除主观估计偏差，直接掌握大量第一手实际资料，且能揭示或确立现象之间的相关关系，在定量分析上具有重要作用。但由于实验对象和实验环境的选择难有充分的代表性，实验调查的结论带有一定的特殊性。且实验法对调查者要求较高，花费时间较长，故其应用范围有限。

实验法一般适用于对新设计、新包装、新价格、新配方、新广告等社会经济现象的实践效果资料搜集。

4. 报告法

报告法是由被调查单位按照调查机关的调查方案要求，及时向调查机关报告统计资料的调查方法。统计报表就属于这种方法。报告法一般是对机关团体和企事业单位，而不是对个人调查，下级必须按规定准确及时、全面地向上级提供统计资料，具有法律行政的强制性。目前在我国，只要求规模在规定标准以上的企业执行统计报表制度。对于规模在规定标准以下的企业及个体经营经济单位，则采用抽样调查的方法，仅仅对抽入样本的单位，按照国家统一制定的统计报表制度来搜集原始数据。

报告法在统计报告系统健全、原始记录和核算工作完整的前提下，可以保证提供资料的准确性。但报告法只能用来获取社会与经济发展重要指标体系所需要的原始数据，不可能覆盖社会经济生活的方方面面。而且通过颁发调查方案来搜集资料，调查者与被调查者不直接接触，不能了解社会经济活动的具体情况和影响原因，在经济利益主体多元化的情况下，容易发生虚报瞒报现象。

报告法常应用于对无法进行直接观察、访问和实验的历史资料的搜集。

5. 文献法

文献法是调查人员根据调查方案的内容和要求，搜集文献资料的一种方法。文献包括报纸、书籍、数据表格等文字、数字文献，也包括影视、图画、磁带、唱片等声音、图像文献。

文献法与实地调查方法相比，调查受控制因素少，花少量人力、财力可获得所需资料。这种方法，不仅适用于社会经济调查，而且广泛运用于科学研究。但文献法对文献资料的真实性的鉴定工作难度大。

以上调查方法中，观察法、访问法、实验法都是直接搜集第一手统计资料的方法，报告法和文献法则是间接搜集第二手统计资料的方法。它们均是传统调查方法，在社会实践中得到了广泛的应用。

2.4.2 网上调查

网上调查(survey on the net)是指利用互联网进行调查，获取调查资料的统计调查方



法。它有两种方式，一种是利用互联网直接进行问卷调查等方式搜集一手资料，这种方式称为网上直接调查；另一种是利用互联网的媒体功能，从互联网搜集二手资料，这种方式一般称为网上间接调查。由于越来越多的报纸杂志、电台等媒体，还有政府机构、企业等也纷纷上网，因此网上成为信息的海洋，信息蕴藏量极其丰富，关键是如何发现和挖掘有价值信息。

1．网上直接调查

网上直接调查一般有以下途径。

（1）设置在线调查表。在网站上设置调查表，访问者在线填写并提交到网站服务器，这是网上调查最基本的形式，广泛地应用于各种调查活动，这实际上也就是问卷调查方法在互联网上的延伸。

（2）利用电子邮件调查。同传统调查中的邮寄调查表的道理一样，将设计好的调查表直接发送到被调查者的邮箱中，或者在电子邮件正文中给出一个网址链接到在线调查表页面。这种方式在一定程度上可以对用户成分加以选择，并节约被访问者的上网时间，如果调查对象选择适当且调查表设计合理，往往可以获得相对较高的问卷回收率。

（3）对网站访问者进行抽样调查。利用一些访问者跟踪软件，按照一定的抽样原则对某些访问者进行调查，类似于传统方式中的拦截调查。

（4）对固定样本调查。同传统调查中的固定样本连续调查法一样，用合理的抽样技术选定固定样本用户，当然，这些用户必须是可以经常上网的用户。对固定样本用户给予必要的培训，说明调查目的，提出一定的要求，由各样本用户按照要求将所要调查的内容记录下来，定期提交给市场调研项目的负责人。

2．网上间接调查

网上间接调查常用以下一些方法。

（1）网上搜索法。利用网上搜索可以收集到所需要的大部分第二手资料。例如，大型调查咨询公司的公开性调查报告，大型企业、商业组织、学术团体、著名报刊等发布的调查资料，政府机构发布的调查统计信息等。

（2）网站跟踪法。网上每天都出现大量的信息，即使功能最强大的搜索引擎，也不可能将所有信息都检索出来，而且很多有价值的信息并不是随便可以检索得到的。作为市场调研的日常资料收集工作，这就需要对一些提供信息的网站进行定期跟踪，对有价值的信息及时收集记录。

（3）加入邮件列表。一些网站为了维持与用户的关系，常常将一些有价值的信息以新闻邮件、电子刊物等形式免费向用户发送，通常只要进行简单的登记即可加入邮件列表。将收到的邮件列表信息定期处理，也是一种行之有效的资料收集方法。

网上调查与传统调查方法相比，具有许多优越性。主要表现在及时性和共享性，便捷性和低费用，无时空、地域限制等。但同时网上调查也存在局限性，主要表现在调查表设计的科学性、样本的数量和质量保证、个人信息保护等问题。

2.4.3 统计调查技巧

实施统计调查，除强调运用科学的方式方法外，还要善于运用调查技巧，这对于提高调查能力和效果尤其重要。

1. 调查问卷设计技巧

调查问卷又叫询问表或调查表，它是指为统计调查所用的、以设问的方式表述问题的表格。调查问卷广泛运用于民意调查和现代市场调查中，是收集资料的最重要工具。

在问卷设计中，为保证问卷的回收率及调查质量，应注意以下几个方面。

（1）问卷中问题的表述要客观、准确、具体，避免抽象、笼统和有多重含义，不能带诱导性和倾向性。例如，“您认为今年市场供应和物价怎样?”这种提问既笼统又同时在一个问题中问两件事，有多重含义，使被调查者难以回答。

（2）对于敏感性问题，不要直接提问，可采用假定法、转移法等。例如，“假定对人口生育不加限制，您认为多子女和独生子女哪种情况更有利于培养子女成材?”这个问题采用了假定法。这样会比直接提问更有利于被调查者回答。

（3）问卷中问题答案的设计既要遵循互斥性原则，又要遵循完备性原则。互斥性原则是指同一问题的若干答案之间关系是相互排斥的，不能有重叠、交叉和包含等情况；完备性原则是指所排列出的答案应包括问题的全部表现，不能有遗漏。

（4）问卷中所设计的一系列问题，要讲究排列顺序。一般先易后难，且符合逻辑性。这样能给被调查者一种轻松、方便、流畅的感觉，以顺利完成调查工作。

（5）问卷设计一定要通过小规模访谈进行修改。

2. 调查谈话技巧

调查谈话是由被调查者回答调查者所提问题而获得调查信息的一种方法。其关键是谈话过程能否顺利完成。要取得成功，调查者应熟练掌握和运用各种技巧。

（1）重视调查谈话的开始。调查者要讲究仪表风度，会寻找共同话题，友好、亲切地接近被调查者，谈话开始的前几分钟对访谈的成败至关重要。

（2）调查提问要简洁，问题要具体，态度要自然。按照由易到难的逻辑顺序进行。

（3）通过采用多种方式提问和巧妙引导的方法，控制谈话过程。

（4）注意利用表情、动作等非语言因素引导被调查者的谈话朝着既定目标发展。

（5）做好谈话的结尾。实际调查中，要审时度势，尽量使谈话在最佳点上结束，不要在分歧中告别，要在笑声中分手。

本 章 小 结

本章共有 4 节，安排两类内容：第一类是统计调查的理论问题，包括统计调查的意义和种类等；第二类是统计调查的实施问题，包括统计调查方案如何设计、调查方式的选择、调查方法和调查技巧的具体运用。



统计调查就是搜集统计资料的工作过程，包括对原始资料和次级资料的搜集。它要求调查所得结果准确、及时、完整、系统，调查所耗费用比较低。统计调查根据调查对象的范围、调查登记时间是否连续、搜集资料的方法不同分为不同的种类。

每项统计调查实施之前，都要设计周密的调查方案，其主要内容包括以下几点：确定调查目的；确定调查对象和调查单位；确定调查项目和调查表；确定调查时间和调查期限；确定调查的组织实施计划。

统计调查的方式主要有以下5种。

（1）普查。它是一种专门组织的、一次性的全面调查。在获得反映国情、国力的基本统计资料方面，普查比其他任何一种调查都具有优越性。

（2）抽样调查。它是按照随机原则从总体中抽出部分单位组成样本，用样本资料推算总体数量特征的一种非全面调查。抽样调查“准确性高、时效性强、经济性好、应用性广”，是国际公认最普遍采用的统计调查方法。

（3）统计报表。它是定期报送国民经济基本统计资料的一种统计报告制度。在计划经济时代被非常充分地应用。目前，统计报表是我国统计调查方法体系的必要补充。

（4）重点调查。它是对重点单位进行的一种非全面调查。重点单位是指它们的数目在全部单位数中只是一小部分，但它们的标志值的总和在总体的标志总量中却占很大的比例。

（5）典型调查。它是在对所研究的现象总体进行初步分析的基础上，有意识地选取有代表性的单位进行的一种非全面调查。

统计调查方法主要包括两种。一是传统调查方法，有观察法、访问法、实验法、报告法和文献法。二是网上调查，有网上直接调查和网上间接调查两种调查方法。统计调查技巧包括调查问卷设计技巧和调查谈话技巧。

应用与拓展

第九次全国国民阅读调查：2011年人均阅读图书4本 超半数国民自认阅读量少

由中国新闻出版研究院组织实施的第九次全国国民阅读调查2012年4月23日正式发布。调查显示，2011年人均阅读图书、报纸和期刊分别为4.35本、100.70期(份)、6.67期(份)，年人均阅读电子书1.42本，18～70周岁国民包括书报刊和数字出版物在内的各种媒介综合阅读率为77.6%，同比增长0.5个百分点。

所谓综合阅读率，是指国民对包括书报刊和数字出版物在内的各种媒介综合阅读状况。但阅读率的提升并没有带来人们对自身阅读状况的肯定。调查显示，50.7%的受访国民自认阅读量少。有1.2%的国民认为自己阅读数量很多，7.6%的国民自认阅读数量比较多，而40.5%的国民认为自己的阅读量一般。

尽管调查中的综合阅读率略有增长，但传统阅读状况却不容乐观，只有图书阅读率同比增加了1.6个百分点，而报纸和期刊阅读率都有不同程度下降。形成对照的是，数字化阅读呈现明显增长势头。2011年，网络在线阅读人数比2010年增加了11.1%，人均每天上网时长为47.53分钟，比2010年的42.73分钟增加了4.8分钟。

调查还发现，有27.6%的被调查者用手机来阅读，比2010年增加了4.6个百分点。目前，我国人均每天手机阅读时长为13.53分钟，比2010年增加了3.21分钟。此外，手机阅读人群的年人均花费已达20.75元。

新闻出版研究院院长郝振省分析说，获取便利是读者选择数字阅读的首要原因，此外，随时随地阅读、方便检索信息也是读者转向数字阅读的两个重要原因，与此相比，传统纸媒不具有信息检索方面的优越性。

价格便宜是数字阅读另一大优势。调查表明，41.8%的人群能够接受付费下载阅读，他们能够接受的电子书平均价格为每本3.5元，而对于购买一本200页左右的文学类纸质简装书，国民能接受的平均价格为13.43元。

（资料来源：http://news.xinhuanet.com/society/2012-04/23/c_111829614.html.）



习题与实训

一、单项选择题

1. 问卷调查、专家访谈、电话调查等方法都属于(　　)。

A. 直接调查　　B. 间接调查
C. 既是直接调查又是间接调查　　D. 以上都不是

2. 对家乐福超市全体员工进行身体健康状况调查，调查单位是(　　)。

A. 每位员工　　B. 所有员工　　C. 所有商品　　D. 每一件商品

3. 普查人口2010年11月1日零时的状况，要求将调查单位的资料在2010年11月10日前登记完成，则普查的标准时间是(　　)。

A. 2010年10月31日24时　　B. 2010年11月10日零时
C. 2010年11月9日24时　　D. 2010年11月1日24时

4. 某市2012年工业企业经济活动成果的统计年报的呈报时间为2013年1月31日，则调查期限为(　　)。

A. 1年　　B. 1年零1个月　　C. 1个月　　D. 1天

5. 调查大庆、胜利、大港、中原的几个大油田，以了解我国石油工业生产的基本情况，这种调查属于(　　)。

A. 普查　　B. 重点调查　　C. 概率抽样调查　　D. 典型调查

6. 为了研究大中型国有企业改革成功的经验和失败的教训，从全部大中型国有企业中选出若干先进的和失败的代表单位组成样本进行调查，比较适合的调查方式是(　　)。

A. 普查　　B. 重点调查　　C. 概率抽样调查　　D. 典型调查

7. 下列情况的统计调查，属于一次性调查的是(　　)。

A. 商品库存量　　B. 商品购进额
C. 商品销售量　　D. 商品销售额

8. 在下列调查中，调查单位与填报单位一致的是(　　)。

A. 公司设备调查　　B. 农村耕地调查
C. 学生学习情况调查　　D. 汽车养护情况调查



9. 要了解糖果包装的改变对销售量的影响情况，则选定一个地区，将新旧两种包装的糖果投入市场进行试验对比，观察其销售量变化和消费者反映，获得数据作为新包装是否采用的依据，这种调查方法是(　　)。

A. 观察法　　B. 实验法　　C. 报告法　　D. 访问法

10. 要调查人群中经常上网浏览的人的年龄、性别、职业等情况，比较适宜的调查方法是(　　)。

A. 观察法　　B. 访问法　　C. 实验法　　D. 网上调查

二、多项选择题

1. 在经济普查中(　　)。

A. 第二、第三产业总的企业数是调查对象

B. 每一个企业是调查单位

C. 每一个企业是填报单位

D. 企业的所有制性质是变量

E. 每个企业的营业额是调查项目

2. 下列调查中，属于一次性调查的有(　　)。

A. 人口普查　　B. 职工家庭收支变化调查

C. 第三产业从业人数调查　　D. VCD库存量调查

E. 公司利润调查

3. 下列情况的统计调查，其调查单位和填报单位不一致的是(　　)。

A. 手机质量调查　　B. 网络用户年龄调查

C. 大学生睡眠时间调查　　D. 居民住宅调查

E. 城市居民生活状况抽样调查

4. 普查属于(　　)。

A. 全面调查　　B. 非全面调查　　C. 经常性调查

D. 一次性调查　　E. 专门调查

5. 下列调查中，属于经常性调查的有(　　)。

A. 企业的产品产量　　B. 公司的利润额

C. 商品库存量　　D. 商业网点数

E. 商品的销售量

6. 统计报表的资料来源有(　　)。

A. 原始记录　　B. 调查问卷

C. 基层单位内部报表　　D. 基层单位统计报表

E. 统计台账

7. 下列调查中属于直接调查搜集第一手资料的方法有(　　)。

A. 观察法　　B. 电话调查法　　C. 实验法

D. 文献法　　E. 网上问卷调查

8. 要检查节日市场上肉制品质量，有关部门采用随机原则抽取部分肉制品进行调查，此项调查属于(　　)。

A. 全面调查　　B. 非全面调查　　C. 典型调查

D. 概率抽样调查　　E. 专门调查

9. 我国第六次人口普查的标准时间是2010年11月1日零时，下列情况应统计人口数的有(　　)。

A. 2010年11月2日出生的婴儿

B. 2010年10月31日出生的婴儿

C. 2010年10月31日晚死亡的人

D. 2010年11月1日1时死亡的人

E. 2010年10月31日出生，11月1日5时死亡的人

10. 下列调查方法属于访问法的是(　　)。

A. 电话调查法　　B. 面谈法　　C. 文献法

D. 邮寄问卷法　　E. 观察法

三、判断题

1. 代表性误差不存在于普查中。(　　)

2. 我国经济普查今后每10年进行两次，因此，它是一种经常性调查方法。(　　)

3. 统计调查方案中要确定调查时间与调查时限，目的是保证统计调查的及时性和完整性。(　　)

4. 要了解“十一”黄金周期间我国铁路旅客周转量，只需对全国几个大的铁路枢纽客运量进行调查，就可以掌握全国基本情况。这种调查属于非全面调查。(　　)

5. 重点调查中的重点单位是根据进行统计调查时，当前工作的重点来确定的。(　　)

6. 典型调查由于在选取典型单位时已对所研究对象进行了全面分析，故可以用典型调查的结果来精确地推断总体。(　　)

7. 在抽样调查中，总体可以是有限总体也可以是无限总体，样本是有限的。(　　)

8. 某地区组织一次物资普查，要求1月10日～27日全部调查完毕，这一时间是调查时限。(　　)

9. 单一调查表就是在一张表上只登记一项调查内容的表格。(　　)

10. 传统统计调查方法都是直接调查第一手统计资料，网上调查可以搜集第二手统计资料。(　　)

四、填空题

1. 世界各国在反映本国综合实力的国情国力调查中，一般都采用________的调查方式来完成。

2. 专门调查主要有________、________、________和________4种形式。

3. 某管理局要求所属企业在2013年1月15日上报2012年年末从业人员资料，则调查时间是________，调查期限是________。

4. 调查人员亲临现场对调查单位进行清点和计量，这种调查方法称为________。

5. 典型调查中的典型单位是________选取的，抽样调查中的样本单位是________选取的。

6. 无论采用何种调查方法进行调查都要先制定________。



7. 调查表一般由________、________和________3部分组成。

8. 由调查人员通过口头、书面等方式向被调查者了解情况，取得第一手统计资料的调查方法称为________。

9. 调查问卷中问题的设计有两种，即________和________。

10. 我国目前的统计调查模式是以________为基础，________为主体，同时辅之以重点调查、科学推算和少量的全面报表综合运用的统计调查方法体系。

五、应用能力训练题

1. 某学院拟对大学生的课外生活进行一次问卷调查。调查项目包括课外时间的分配和利用，课外活动的形势及时间占用，对大学生勤工俭学的看法和建议。试设计简单的统计调查方案。

2. 指出表2-1中调查的调查对象、调查单位和填报单位。

表2-1 与调查相关的训练题

调查内容	调查对象	调查单位	填报单位
手机质量调查			
网络用户年龄调查			
超市商品物价调查			
居民住宅调查			
高等院校调查			
大学生睡眠时间调查			
城市居民生活状况抽样调查			

3. 从统计调查对象包括的范围、调查登记时间是否连续、搜集资料的方法是否相同等方面对以下统计调查实例分类，并指出各属于哪种统计调查方式。

(1) 2013年，对我国的工业企业从业人数进行调查，各企业按上级部门要求填报统计表。

(2) 2013年，对全国所有第二、第三产业活动单位进行基本情况摸底调查，以2013年12月31日为标准时点，调查2013年度的资料。

(3) 对进口的一批产品，抽检其中的少部分以对整批产品质量进行评价。

(4) 要了解全国钢铁产量的基本情况，只要对全国少数几个重点钢铁企业如鞍钢、宝钢、首钢、武钢等进行调查，就能及时对全国钢铁产量的基本情况进行推断。

(5) 为了探讨一项新改革措施实施的效果，推广其成功经验，对已采取改革措施并效果明显的代表性单位进行调查。

4. 指出表2-2中调查的统计调查方法，它们属于直接调查还是间接调查。

5. 以你所在的班级为总体进行统计调查，总体单位是每一位同学，调查的有关标志是学生的身高、体重、性别和年龄。不出现姓名。

(1) 请设计一个简单的统计调查方案。

表 2-2　辨析调查方法训练题

调 查 描 述	统计调查方法	直接调查或间接调查
为获得某高速路口汽车交通量资料，调查人员亲自在路口对出入口各种汽车数量进行观察、记数		
食品厂将新旧两种配方的面包，委托食品店调查各自销售量，以决定是否更改配方		
经过周密设计，对抽选出的 1 000 位市民发放调查问卷，进行“社会热点问题看法”专题调查		
为了解同学们使用计算机的习惯，在网上设计问卷进行“校园 E 族调查”		

(2) 设计一个单一调查表，包括表头、表体和表脚。

每个同学将各自标志的结果写在纸条上，搜集在一起，就形成了原始资料。请保存好这些资料，在后面的章节中我们会利用这些数据。

6. 你所在的学院伙食如何？大二学生每月零花钱数额为多少、来源及去向是怎样的？对这些问题或者你感兴趣的其他问题，进行模拟统计调查。尝试设计一份调查问卷，注意利用调查问卷设计技巧。



第3章 统计整理

理论目标

(1) 理解统计整理的意义及步骤。
(2) 掌握统计分组的基本理论和方法。
(3) 掌握分配数列的特性与编制方法。
(4) 了解统计图表结构、种类和编绘方法。

能力目标

(1) 能将资料整理成分配数列。
(2) 能根据分配数列绘制统计图表。

做一做

李华大学毕业后与人合伙开了一家超市，超市在风雨中走过了3年，2011年达到销售收入990万元，为了了解本地区超市总体经营状况，同时可判断自己经营的超市在本地所处的地位，他搜集了50家超市商品销售额的数据(单位：百万元)，如下所示。

7.4	12.6	29.0	2.0	12.4	7.0	14.8	17.5	15.0	18.2
18.7	15.5	12.8	26.0	17.3	8.3	14.7	12.0	3.5	6.8
25.0	19.3	6.4	4.0	11.9	8.5	13.2	14.5	17.1	15.6
13.4	4.5	9.5	20.0	15.7	6.0	11.4	23.0	14.2	16.7
21.0	16.0	13.6	10.0	13.9	5.0	5.8	10.5	16.3	22.0

他需要把这些资料分组整理并形成一个简短报告，便于与合伙人共同分析。请你帮李华做一做。

■ **想一想**

现实当中搜集的资料大多是杂乱无章的原始资料，要想对它们进行分析，必须先经过资料的整理环节。这就像蔬菜买回家不可直接烹饪，必须经过择菜、清洗等步骤一样。对调查得到的数据资料应该按什么步骤整理？每一步应该怎样正确操作？如何经过数据的处理，使原本繁杂的数字变为能反映总体综合特征的系统化、条理化的资料，而且可以图文并茂、生动直观地表现出来？本章将按照统计整理的步骤，逐一进行介绍。

3.1　统计整理的意义和步骤

3.1.1　统计整理的意义

统计整理(statistical treatment)指根据统计研究的任务和要求，对统计调查所取得的原始资料进行科学的分类、汇总或对已整理过的资料进行再加工，使之成为系统化、条理化，能反映现象总体特征的综合资料的工作过程。

统计整理是统计工作中不可或缺的重要环节，具有承前启后的作用。统计整理不仅是一项简单的汇总工作，它还是统计调查的继续，也是统计分析的前提。如果没有统计整理，即使统计调查所得到的资料非常丰富、完善，其作用也难以发挥，统计工作也就无法继续。统计调查所搜集到的大量的反映个体特点的资料，只有通过科学的整理加工，形成反映总体特征的综合资料，才能为下一步的统计分析工作做好准备。

3.1.2　统计整理的步骤

统计整理是一项细致、科学的工作。一项完整的统计资料整理工作包括拟订整理纲要、审核、分组汇总、编制统计图表及统计资料的保管与积累5个方面。其中，核心内容是对统计资料的审核、分组、汇总和编制统计图表等环节。

1. 审核原始资料

在搜集资料的过程中，经常会由于某些原因出现一些差错。因此，为了确保统计资料准确无误，以及符合统计研究目的的要求，必须对统计调查所获得的原始资料进行严格的审核，发现问题及时纠正。审核是统计整理中的一个重要的环节，主要包括对资料准确性、及时性和完整性的审核。

审核资料的准确性，就是检查所填报的资料是否准确和可靠。审核主要从两个方面进行：一是逻辑检查，主要从理论上或常识上检查资料的内容是否有悖常理、有无不切实际或不符合逻辑的地方，以及各项目之间有无相互矛盾之处；二是计算检查，主要检查填报单位有无遗漏、调查表项目是否填写完整、所填内容和表格规定是否一致、计量单位与法定单位是否一致，以及各项数字之间的关系是否正确等。

审核资料的及时性，主要检查资料是否符合调查规定的时间，是否在规定的调查期限报出。



审核资料的完整性，主要检查调查单位资料是否齐全，是否上报了规定的份数、调查项目是否完整等。

2. 进行分组汇总

分组汇总就是用一定的组织形式和方法对经过审核的资料进行分组、汇总和计算。根据研究目的和统计分析的需要，选择整理的标志，并进行划类分组。科学的分组是做好统计整理的前提条件，只有正确地分组才能整理出有科学价值的综合指标，并借助这些指标揭示现象的本质与规律。

3. 编制统计图表

根据统计分析的要求或社会经济现象之间的联系，对整理好的资料编制成统计表或绘制成统计图，简明扼要并且形象直观地反映社会经济现象之间的数量特征。

3.2 统计分组

3.2.1 统计分组的概念和作用

统计分组(statistical classification)是根据统计研究的需要和总体的内在特征，将总体按照一定的标志划分为若干个组成部分的一种统计方法。简而言之，统计分组就是把性质相同的总体单位归为一组，不同性质的总体单位区别开来形成不同的组。社会现象是复杂的，现象之间既有某种共同的性质，同时又存在差异，通过统计分组，揭示现象各部分之间存在的差异，结合统计分析，能够认识事物的发展变化规律。例如，企业按资产大小划分为大型企业、中型企业和小型企业；我国将国民经济产业结构划分为第一产业、第二产业和第三产业；将考试成绩划分为不及格、及格、中、良好、优秀5个等级等。统计分组的结果是保持组内资料的一致性和组间资料的差异性。

理解统计分组应从两个方面进行：一方面，对总体而言是“分”，即将总体按某个标志划分为若干个性质相异的组成部分；另一方面，对总体单位而言是“合”，即将具有某种相同性质的总体单位归在同一组。

统计分组是统计整理的关键，只有对总体进行科学的分组，才能对统计资料进行进一步的科学的整理和分析以得到有价值的结论。

统计分组是统计研究中应用的基本方法之一，贯穿于统计研究的整个过程，在统计研究中占有非常重要的地位。统计分组的作用主要体现在以下3个方面。

1. 划分现象的类型

划分现象的类型是统计分组的主要作用。社会经济现象千差万别，存在着不同的类型，而且各种不同的类型具有不同的特点和发展规律。要了解各种社会经济现象的性质、特点、相互关系及变化规律，必须运用统计分组将现象总体划分为性质不同的类型，以便揭示不同社会经济现象的质的差异。例如，根据学历教育类型进行划分时，我国各级各类

学历教育学生的情况见表3-1。

表3-1 2010年我国各级各类学历教育学生情况

学历教育类型	在校生数/人	学历教育类型	在校生数/人
高等教育	34 168 470	工读学校	10 735
中等教育	100 129 290	特殊教育	425 613
初等教育	101 353 616	学前教育	29 766 695

(资料来源:中华人民共和国教育部2010年教育统计数据.)

通过这种分类,可以反映我国学历教育学生在不同类型分布的情况,也为进一步研究我国学历教育学生结构提供了依据。

2. 反映现象总体的内部结构

从数量上反映总体的内部结构是统计研究的重要任务。社会现象包含大量的单位,由于它们在性质上不尽相同,而且在各种类型总体中的比例也不同,因而所处的地位和对总体的影响程度也不同,不同的类型分布决定了不同的总体性质。将社会经济现象总体经过统计分组后被划分为不同性质的组成部分,计算总体内部各组成部分占总体的比例,分析总体各部分的性质、结构和比例关系。另外,还可以将总体内部结构分组资料按时间的发展顺序进行对比分析,从而认识现象发展变化的规律。例如,2005—2011年,我国按产业分类的从业人员的构成情况见表3-2。

表3-2 2005—2011年我国按产业分类的从业人员构成情况

年　份	第一产业	第二产业	第三产业	合　计
2005	44.8%	23.8%	31.4%	100%
2006	42.6%	25.2%	32.2%	100%
2007	40.8%	26.8%	32.4%	100%
2008	39.6%	27.2%	33.2%	100%
2009	38.1%	27.8%	34.1%	100%
2010	36.7%	28.7%	34.6%	100%
2011	34.8%	29.5%	35.7%	100%

从表3-2中可以看出,2005—2011年这7年间我国在3类产业中的从业人员发生了变化,第一产业从业人员逐渐下降,第二产业和第三产业从业人员有所上升。

3. 分析现象之间的依存关系

统计研究的社会经济现象不是彼此孤立的,总是存在相互依存和相互制约的关系。例如,国民收入与居民储蓄额之间、市场商品价格与其需求量之间、商品销售额和广告费用之间、家庭的工资收入与生活费支出之间、工人技术级别与产品质量之间、工人劳动生产



率与产品成本之间，等等，都在一定程度上存在相互依存的关系。这些依存关系，可通过统计分组来发现影响因素与结果因素之间的变动规律，并可进一步从数量上描述依存关系的密切程度。例如，我国民营企业规模与品牌竞争力的关系见表3-3。

表3-3 我国民营企业规模与品牌竞争力的关系

企业规模/人	品牌竞争力均值（PPC）	企业数/家	企业规模/人	品牌竞争力均值（PPC）	企业数/家
100以下	17.249 3	15	500～1 000	24.514 9	44
100～200	19.593 9	28	1 000～2 000	28.913 2	186
200～500	21.126 2	10	2 000以上	44.990 2	107

（资料来源：中国社会科学院民营经济研究中心，2005年中国制造业民营企业品牌竞争力指数500家数据库.）

从表3-3中可以看出，企业规模按员工人数分为不同的级别，随着民营企业规模的不断扩大，其对应的品牌竞争力指数也在增长。这说明我国民营企业规模的发展成长，使品牌竞争优势增强，规模与竞争力之间是存在相互依存关系的。

3.2.2 统计分组的种类

1. 按分组的作用和目的划分

按分组的作用和目的不同，统计分组可分为类型分组、结构分组和分析分组。

（1）类型分组是将复杂的现象总体，划分为若干个不同性质的部分。

（2）结构分组是在对总体分组的基础上计算出各组占总体的比重，以此研究总体各部分的结构。

（3）分析分组是为研究现象之间的依存关系而进行的统计分组。

一般情况下，类型分组和结构分组总是紧密联系在一起的。

2. 按分组标志的多少划分

按分组标志的多少及其排列形式，统计分组可分为简单分组、复合分组和分组体系。

（1）简单分组也称为单一分组，就是对总体只按一个标志进行分组。例如，上网计算机按上网方式划分为拨号上网、专线上网、ISDN上网、宽带上网和其他方式上网等；企业按生产规模可分为大型、中型和小型3组；货运按运输方式可分为铁路运输、公路运输、水陆运输、航空运输与管道运输5组。

（2）复合分组是对所研究的总体按两个或两个以上的标志进行的多层次分组，即在按某一标志分组的基础上再按另一标志进一步分组。例如，人口在按性别分组的基础上再按年龄分组，企业在按经济类型分组的基础上再按规模进行分组。再如，企业先按所有制性质分组，然后再按规模分组，见表3-4。

采用复合分组方式，可以从同一现象的层层分组和分组标志的联系中，更深入地反映总体的内部结构，更细致、全面地分析问题。当采用一个分组标志不能充分说明现象或现象受两个或两个以上因素影响时，选用复合分组比简单分组更能清晰地反映现象的结构。

但是，复合分组的组数会随着分组标志的增加而成倍增加，使每组包括的单位数相应减少，处理不好就会影响分析问题。可见，在采用复合分组时，分组标志选择要适量，而且总体包括的单位数应较多。究竟采用几个分组标志进行分组，要根据统计研究的目的和任务决定。

表 3-4　某地区某年年底各类工业企业及规模构成情况

按所有制性质及规模分组	企业数/家	比　　重
全民所有制企业	3 204	64.08%
大　型	512	10.24%
中　型	708	14.16%
小　型	1 984	39.68%
集体所有制企业	1 286	25.72%
大　型	286	5.72%
中　型	434	8.68%
小　型	566	11.32%
私营企业	152	3.04%
大　型	24	0.48%
中　型	48	0.96%
小　型	80	1.60%
中外合资企业	212	4.24%
大　型	70	1.40%
中　型	74	1.48%
小　型	68	1.36%
外商独资企业	102	2.04%
大　型	28	0.56%
中　型	36	0.72%
小　型	38	0.76%
其他类型企业	44	0.88%
大　型	6	0.12%
中　型	10	0.20%
小　型	28	0.56%
合　　计	5 000	100.00%

(3) 分组体系是采用一系列相互联系、相互补充的并列标志，对被研究对象总体进行的分组。社会现象是复杂的，无论简单分组还是复合分组，都只能对现象从一个或几个方面进行观察与分析。为获得对复杂现象全面的认识，应采用分组体系从不同角度、不同方面对某一经济现象进行观察和分析研究。例如，研究我国网民状况时，可以按性别、婚姻状况、年龄、受教育程度、个人月收入、职业、行业等多种标志进行分组，这些标志互相联系、互相补充，充分说明了我国网民的构成状况。

3. 按分组标志的性质划分

按分组标志的性质，统计分组可分为品质标志分组和数量标志分组。



(1) 品质标志分组是按品质标志进行的分组，即按事物的某种属性分组。品质标志分组可以反映总体内部的性质差异。例如，企业按经济类型分组；人口按性别、民族、受教育程度、职业分组等。

(2) 数量标志分组是按数量标志进行的分组。例如，人口按年龄分组、企业按固定资产价值分组、网民按个人月收入分组等。数量标志分组可以反映现象总体数量上的差异。

3.2.3 分组标志的选择及界限的确定

统计整理的关键在于统计分组，统计分组的关键在于正确选择分组标志和划分各组界限。分组标志是统计分组的依据或标准，正确地选择分组标志能使统计分组的作用得以充分发挥。

1. 分组标志的选择

反映总体单位的标志有许多，标志的选择将影响统计分析的结论的正确性与真实性。正确选择分组标志，必须遵循以下原则。

(1) 应根据研究的目的与任务选择分组标志。统计研究的目的是统计分组标志选择的依据。统计分组是为统计研究服务的，统计研究的目的不同，对分组的要求也不同，选择的分组标志也应有所不同。例如，要分析我国网民所在行业的结构分布，就要选择行业作为分组标志；要分析网民的教育结构，则要选择受教育程度作为分组标志；要分析性别结构，则要选择性别作为分组标志；要分析网民的收入结构，则要选择网民个人月收入作为分组标志。

(2) 选择最能体现现象本质特征的标志作为分组标志。研究某一问题可能涉及许多标志，有些标志是主要标志，有些标志是次要的。主要标志能反映现象的本质。在统计分组时，要选择最能体现现象本质特征的主要标志作为分组标志。例如，研究企业的经济效益时，可供选择的分组标志有总产值、净产值、销售收入、利税额、单位产品成本等，而其中最能反映企业经济效益好坏的标志是利税额，故应选择这一标志作为分组标志。

(3) 结合现象发展的具体历史条件和经济条件选择分组标志。社会经济现象随着时间、地点、条件的变化而不断变化，反映现象本质特征的主要标志也会因时、因地而不同。选择分组标志时，应根据历史的发展“与时俱进”，具体情况具体分析，客观选择分组标志。例如，改革开放以前，当时的“三转一响”，即手表、自行车、缝纫机和收音机就是引领时尚的奢侈品。因此在反映居民家庭生活状况时，应选择家庭拥有手表(自行车、缝纫机、收音机等)数作为主要分组标志；而随着社会的进步和人民生活水平的提高，当年的“三转一响是大件”的情形已经成为历史的记忆。因此，为了真实反映居民家庭生活状况，需要对分组标志进行调整，20 世纪 80 年代可以选择拥有彩电、冰箱、洗衣机数等作为分组标志；目前则常以家庭拥有房产、轿车数等作为反映居民家庭生活状况的主要分组标志。

2. 分组界限的确定

确定分组标志之后，还必须划分各组界限。划分各组界限是在分组标志的变异范围

内，划定各相邻组间的性质界限和数量界限。划分各组界限，一定要本着保证各组组内单位的同质性和组与组之间单位的差异性的原则进行。

总之，分组标志的选择与分组界限的划分是统计研究中非常关键的一步，必须遵守相关的原则认真分析，做出正确的判断，选出合理的分组标志，为进一步的统计分析做好准备。

3.2.4 统计分组的方法

按分组标志的性质不同，统计分组分为品质标志分组和数量标志分组，这种分组类型是最重要的统计分组类型，这两种分组的具体处理方法不同，统计分组方法主要围绕这两种分组进行阐述。

1. 品质标志分组的方法

品质标志分组是指选择反映事物属性差异的品质标志作为分组标志后，根据其变异范围划定各组界限，将总体划分为若干个形式不同的组成部分。

按品质标志分组时，有些情况比较简单，有些情况比较复杂。具体来说，其组数的确定主要取决于两个因素——事物的特点与统计研究的任务。事物本身所具有的既定的属性，是确定组数的基本依据。进行品质分组，其组数的多少首先取决于事物本身的特点。例如，人口按性别分组，就只能分为两组；高等学校在校生按学历可分为专科、学士、硕士和博士 4 组。

对于有些事物构成比较复杂、组数可多可少的情况，就需要根据统计研究的目的和要求确定组数。例如，就业人员按行业进行分组，组数可多可少的，究竟分为几组合适，只能根据统计研究的目的和要求确定了。要求详细时，组数可多些；要求粗略时，组数则可少些。

在统计实践工作中，常常需要对所研究的现象进行复杂的品质分组。对于比较复杂的品质标志分组，为了避免因认识不同而造成的混乱，保证分组的同一性和可比性，国家统计部门制定了统一的分类目录和标准，作为分组的标准和依据，如《统计上大中小微企业划分标准》、《关于划分企业登记注册类型的规定》、《统计用产品分类目录》等，各地区、各部门进行统计资料整理时必须遵照执行。

2. 数量标志分组的方法

数量标志分组是指选择反映事物数量差异的数量标志作为分组标志后，根据其变异范围划定各组界限，将总体划分为若干个性质不同的组成部分。

数量标志分组，不是简单地确定各组间的数量差异，而是要通过分组体现数量变化以确定现象的不同性质和类型。根据变量值的大小来准确划分性质不同的各组界限时，首先要分析总体中有多少种性质不同的组成部分，再确定各组成部分的数量界限。

按数量标志分组的方法有两种：单项式分组和组距式分组。

(1) 单项式分组。是按每个变量值对现象总体进行的分组。单项式分组适用于离散变量并且变量值的变动范围不大的情况。例如，某小区居民家庭拥有汽车的情况见表 3－5。



表 3-5 某小区居民家庭拥有汽车的情况

居民家庭汽车拥有数量/辆	居民户数/户	居民家庭汽车拥有数量/辆	居民户数/户
0	68	3 及以上	2
1	24	合　计	100
2	6		

（2）组距式分组。是按变量值的一定范围对现象总体进行的分组。将现象的总体变动范围划分为若干个区间，以每个区间作为变量值。组距式分组适用于连续变量和变动范围较大的离散变量的情况。例如，将 500 只灯泡按耐用时间分组见表 3-6。

表 3-6 500 只灯泡按耐用时间分组

耐用时间/小时	灯泡数/只	所占比重
800～850	35	7%
850～900	127	25.4%
900～950	185	37%
950～1 000	103	20.6%
1 000～1 050	42	8.4%
1 050～1 100	8	1.6%
合　计	500	100%

3.3 分配数列

3.3.1 分配数列的意义和种类

在统计分组的基础上，将总体的所有单位按组进行归类整理并按一定顺序排列，计算出各组的单位数，形成一个反映总体中各单位在各组中的分布情况的数列，这个数列称为分配数列或次数分布(frequency distribution)。

分配数列是统计整理的一种重要形式，也是进行统计描述和统计分析的基础。它可以反映总体的结构分布状况和分布特征，并且可以由此分析研究总体中某一标志的平均水平及其变动规律。

分配数列由两个要素构成：一是总体中按某标志分的组，二是各组相应的分配次数或频率或标志值。在分配数列中，分布在各组的总体单位数称为次数，又称为频数(frequency)；各组次数占总体次数的比重，称为比率，又称为频率。次数和频率从不同角度反映了各组标志值出现的频繁程度，说明了总体单位在各组中的分布，是分配数列的两种表现形式。

分配数列按分组标志特征的不同，分为品质分配数列和变量分配数列。

1. 品质分配数列

品质分配数列简称品质数列，是按品质标志分组形成的分配数列。它在品质分组的基础上形成，主要用于研究总体构成情况。它由各组名称和次数构成，各组次数以绝对数表示，即频数；以相对数表示，即频率。例如，将中国体育代表团在2008年北京奥运会上获得的51枚金牌按金牌项目分组，形成的品质分配数列见表3-7。

表3-7　中国体育代表团在2008年北京奥运会上获得金牌的项目

获得金牌的项目	金牌数/枚	比　重	获得金牌的项目	金牌数/枚	比　重
射箭	1	1.96%	摔跤	1	1.96%
拳击	2	3.92%	击剑	1	1.96%
羽毛球	3	5.88%	游泳	1	1.96%
皮划艇	1	1.96%	竞技体操	9	17.65%
举重	8	15.69%	蹦床	2	3.92%
赛艇	1	1.96%	乒乓球	4	7.84%
帆船	1	1.96%	跆拳道	1	1.96%
跳水	7	13.73%	柔道	3	5.88%
射击	5	9.80%	合计	51	100%

从表3-7中可以看出，该品质数列反映了我国体育代表团在竞技体操、举重、跳水等项目上金牌数目较多。只要品质标志分组正确，形成的分配数列通常都能准确地反映总体的分布特征。

2. 变量分配数列

变量分配数列简称变量数列，是按数量标志分组形成的分配数列。它在变量分组的基础上形成，主要用于反映不同变量值各组的分布情况。它由变量值和次数构成。变量数列按其分组方法的不同，可以分为单项式变量数列和组距式变量数列。

(1) 单项式变量数列。单项式变量数列简称单项数列，是按单项式分组形成的变量数列，每个变量值是一个组，顺序排列。在单项式变量数列中，有多少不重复的变量值就有多少组。单项式变量数列仅适用于变动幅度比较小、不重复变量值较少的离散变量分组的情况。例如，表3-5中的数据为单项式变量数列。

(2) 组距式变量数列。组距式变量数列简称组距数列，是由组距式分组形成的变量数列。每个组由若干个变量值形成的区间表示。组距式变量数列适用于连续变量分组或变动幅度较大、不重复值较多的离散变量分组的情况。例如，表3-6中的数据为组距式变量数列。



3.3.2 变量数列的编制

1. 变量数列次数分布表的编制

变量数列次数分布表的编制，也分为单项式变量数列和组距式变量数列两种情况。

对于总体单位数不多的离散变量，可以直接编制单项式变量数列次数分布表。具体编制方法：按变量值分组列于左方，然后将各变量值出现的次数列于右方，见表3-5。

而对于连续变量或变动幅度较大、不重复值较多的离散变量，则要编制组距式变量数列。组距式变量数列次数分布表编制比较复杂，下面就其编制方法进行详细讲解。

（1）原始数据按数值大小依次排序。

【实例3.1】 某高校经济管理系2013年某班40名学生英语成绩，数据(单位：分)如下。

69 61 75 60 80 72 82 86 90 78
62 87 71 77 79 87 85 97 85 63
94 78 65 85 96 88 89 82 87 78
95 83 56 81 58 75 78 74 67 75

将上述数据(单位：分)进行排序，结果如下。

56 58 60 61 62 63 65 67 69 71
72 74 75 75 75 77 78 78 78 78
79 80 81 82 82 83 85 85 85 86
87 87 87 88 89 90 94 95 96 97

（2）确定组距(class interval)、组数和组限。编制组距式变量数列时，必须要确定组距和组数。

首先应对标志值的分布情况进行仔细观察，找出最大值与最小值并计算出它们的距离(即全距)，确定大多数标量值分布在什么范围，然后确定组距和组数。组距是每个组上限与下限之间的距离，其计算公式为

$$组距=上限-下限$$

组限是组与组之间的界限。每个组的最大值称为上限(upper limit)，每个组的最小值称为下限(lower limit)。将上、下限都齐全的组称为闭口组；将有上限缺下限或有下限缺上限的组称为开口组。开口组一般使用“以上”或“以下”表示。开口组组距以相邻组组距作为参照。

组距与组数关系密切，同一现象的变量分组，组距越大则组数越少，组距越小则组数越多。确定组距与组数应遵循下列原则：尽可能反映总体单位的分布情况及总体单位的集中趋势，体现组内资料的同质性和组与组之间资料的差异性。

组距式变量数列根据组距是否相等，分为等距变量数列和异距变量数列。对于等距变量数列的编制，一般依据总体内部情况的定性分析确定组数，然后用全距除以组数来确定组距，并由此划分各组界限。设 R 为全距，K 为组数，i 为等距组，则有

$$i=R/K$$

在社会实践中，也可参考美国统计学家斯特吉斯(Sturgis)提出的经验公式，即

$$组数=1+3.322\log N$$

根据实际需要而定。

在实例3.1中，最小值为56，最大值为97，全距=97－56=41。根据考试成绩的性质，将组距确定为10，组数为5。

确定组数和组距之后，还要确定组限。组限的确定要遵循这一原则：按这样的组限分组后，标志值在各组的变动能反映总体单位分布的规律性。确定组限时需注意：特殊的界限点必须作为组限(如反映水温的0℃和100℃，反映计划完成程度的100%，反映学习成绩的60分和100分等)。最小组的下限要略低于最小变量值，最大组的上限要略高于最大变量值。

连续变量分组，相邻两组的上限与下限通常是相同的，每一组的上限同时是下一组的下限。这种分组又称为同限分组。为了避免计算时总体单位数出现错误，统计各组单位数时应将到达上限值的单位数计入下一组内，即遵循所谓的“上组限不在内”的原则。例如，成绩刚好为60分的计入60～70分这一组，见表3-9。

离散变量分组，相邻两组的上限与下限可以相同，也可以不相同。相邻两组的上限与下限相同时也遵循“上组限不在内”的原则。相邻两组的上限与下限不相同时，又称为异限分组。例如，工业企业按职工人数分组，见表3-8。

表3-8　离散变量分组的两种表示方法　　　　单位：人

相邻两组上限与下限相同	相邻两组上限与下限不同
100以下	99以下
100～500	100～499
500～1 000	500～999
1 000以上	1 000以上

(3) 编制组距式变量数列次数分布表。确定了组距、组数和组限，经过整理就可绘制组距式变量数列表，见表3-9。

表3-9　40名学生英语成绩分布

考试成绩/分	人数/人	比　重	考试成绩/分	人数/人	比　重
50～60	2	5.0%	80～90	14	35%
60～70	7	17.5%	90～100	5	12.5%
70～80	12	30%	合计	40	100.0%

组距变量数列是以变量的区间进行分组的，掩盖了各组总体单位的实际变量值。为了反映每个组的一般水平，一般采用组中值(mid-value of class)表示。组中值是各组变量的中间数值。组中值的计算公式分开口组与闭口组两类：闭口组的组中值包括同限分组组中值和异限组分组中值两种；开口组的组中值包括缺上限的开口组组中值和缺下限的开口组组中值两种。



同限分组组中值的计算公式为

$$组中值=\frac{上限+下限}{2}$$

异限分组组中值的计算公式为

$$组中值=\frac{本组下限+后一组下限}{2}$$

缺上限的开口组组中值的计算公式为

$$组中值=下限+\frac{邻组组距}{2}$$

缺下限的开口组组中值的计算公式为

$$组中值=上限-\frac{邻组组距}{2}$$

【实例3.2】 计算表3-9中各组数据的组中值，结果见表3-10。

表3-10 某厂工人工资水平组中值计算结果

月工资/元	人数/人	组中值	月工资/元	人数/人	组中值
1 000以下	80	900	1 800～2 000	140	1 900
1 000～1 200	160	1 100	2 000以上	110	2 100
1 200～1 500	270	1 350	合计	1 000	—
1 500～1 800	240	1 650			

第一组的组中值为

$$1\,000-\frac{1\,200-1\,000}{2}=900$$

第二组的组中值为

$$\frac{1\,200+1\,000}{2}=1\,100$$

第三、第四、第五组的组中值的计算类似于第二组的计算，结果见表3-10。

最后一组的组中值为

$$2\,000+\frac{2\,000-1\,800}{2}=2\,100$$

2. 变量数列次数分布图的绘制

一般情况下，我们不仅用次数分布表表示总体单位的分布状况，还可以通过绘制变量数列次数分布图直观、形象地揭示总体单位的分布状况和规律性。

单项式变量数列次数分布图的绘制比较简单。具体绘制方法为：直接以变量值为横轴，以次数为纵轴，在坐标轴上绘出各组变量值与次数相对应的坐标点，然后连接各坐标点即可得到相应的分布图。

组距式变量数列次数分布图一般有直方图(histogram)和次数分配曲线图，但是，次数分配曲线图也是在直方图的基础上绘制的。绘制组距式变量数列次数分布图时，需注意等距变量数列与异距变量数列的处理方法不同。等距变量数列的各组单位数只受变量变化的影响，各组单位数可以直接比较，分组后直接统计整理即可绘制等距变量数列次数分布图。例如，利用表 3－9 中的资料所绘制的等距变量数列次数分布，如图 3.1 所示。

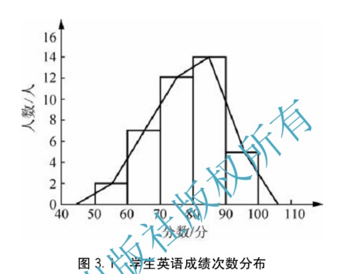


图 3.1 学生英语成绩次数分布

绘制等距变量数列次数分布图的具体步骤如下：以横轴代表变量值，在横轴上标出各组组限值所在位置，以纵轴代表次数并在上面标出各组次数所在位置；然后以组距为宽，以次数为高，绘制相对应的矩形，各组矩形组合构成变量数列总体次数分布特征的直方图。若需要绘制次数曲线分配图，只要在直方图的基础上，连接各矩形顶边的中点(即各组的组中值)，形成一条折线，并且在直方图的左右两端假设各有一个组，将折线与两个假设组的中点连接，就形成了次数分配曲线。这时曲线所覆盖的全部面积与直方图的面积相等。

异距变量数列，由于各组单位数受到组距大小不等和变量值这两种因素的影响，因此需要计算次数密度以消除不等组距对各组次数的影响，进而准确地反映实际的次数分布。次数密度也称为频数密度，即单位组距内分布的次数，其计算公式为

$$\text{次数密度}=\frac{\text{各组次数}}{\text{各组组距}}$$

通过计算次数密度，将不等组距的次数换算为标准组距次数。标准组距为异距变量数列中的各组组距中最小的组距。标准组距次数的计算公式为

$$\text{某组标准组距次数}=\text{该组次数密度}\times\text{该组标准组距}$$

绘制异距变量数列次数分布图的具体步骤与等距数列类似，不同之处在于：以纵轴代表各组标准组距次数并在上面标出各组标准组距次数所在位置；然后以组距为宽，以标准组距次数为高，绘制相对应的矩形，各组矩形组合构成直方图。

【实例 3.3】 根据某厂工人工资水平绘制变量数列次数分布表和次数分布图，分别见表 3－11 和如图 3.2 所示。

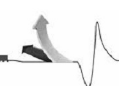


表3-11 某厂工人工资水平次数分布

月工资/元	人数/人	组 距	次数密度	标准组距人数/人
800～1 000	80	200	0.4	80
1 000～1 200	160	200	0.8	160
1 200～1 500	270	300	0.9	180
1 500～1 800	240	300	0.8	160
1 800～2 000	140	200	0.7	140
2 000～2 500	110	500	0.22	44
合计	1 000	—	—	—

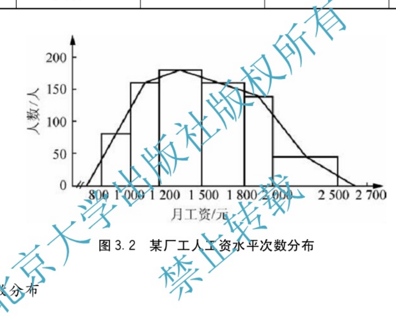


图3.2 某厂工人工资水平次数分布

3. 累计次数分布

借助于次数分布表和分布图可以观察分配数列中各组的次数，以及总体单位数的分布特征。但是若要获取截至某一组变量值以上或以下的分布次数或事物发展进程等情况，就需要编制累计次数分布表。例如，在表3-9中，要计算每个班不及格的学生人数、大于80分的学生人数等数据。

编制累计次数分布表或绘制累计次数分布图都需要计算累计次数或累计频率。累计次数与累计频率分别表明总体的某一标志值在某一水平上下的总体次数与比率。累计次数的计算方法有两种：向上累计和向下累计。

（1）向上累计。向上累计也称为较小制累计，是将各组次数和比率从变量值低的组向变量值高的组逐组累计。组距数列向上累计的次数表明各组上限以下总共包含的总体次数和比率。

（2）向下累计。向下累计也称为较大制累计，是将各组次数和比率从变量值高的组向变量值低的组逐组累计。组距数列向下累计的次数表明各组下限以上总共包含的总体次数和比率。

例如，以表3-9中的40名学生英语成绩分布计算的累计次数分布见表3-12，绘制的累计次数分布图如图3.3所示。

表 3-12　40 名学生英语成绩累计次数分布

考试成绩/分	成绩分布		向上累计		向下累计	
	人数/人	比重	频数	频率	频数	频率
50～60	2	5.0%	2	5.0%	40	100.0%
60～70	7	17.5%	9	22.5%	38	95%
70～80	12	30%	21	52.5%	31	77.5%
80～90	14	35%	35	87.5%	19	47.5%
90～100	5	12.5%	40	100.0%	5	12.5%
合计	40	100.0%	—	—	—	—

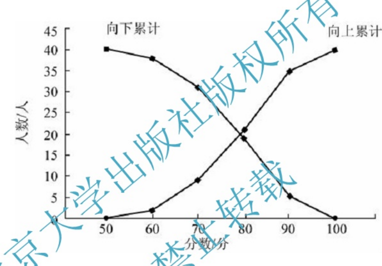


图 3.3　40 名学生英语成绩累计次数分布

从累计次数分布表和累计次数分布曲线图中，可以发现累计次数的特点：同一数值的向上累计和向下累计次数之和等于总体总次数，累计比率之和等于 100%或 1。累计次数分布还是计算位置平均数的依据。

综上所述，当要对大量的、杂乱无章的原始数据进行整理时，首先按某种标志进行分组，然后汇总出每组的单位个数，在这样形成的分配数列的基础上稍作处理，画出图像或进行累计，一方面可以得到反映总体分布特征的直观图像，另一方面还可换个角度表述总体数据的分布规律。例如，针对“做一做”中李华搜集的 50 家超市商品的销售额数据(单位：百万元)，可以按照销售额分组，首先将数据排序，结果如下。

2.0	3.5	4.0	4.5	5.0	5.8	6.0	6.4	6.8	7.0
7.4	8.3	8.5	9.5	10.0	10.5	11.4	11.9	12.0	12.4
12.6	12.8	13.2	13.4	13.6	13.9	14.2	14.5	14.7	14.8
15.0	15.5	15.6	15.7	16.0	16.3	16.7	17.1	17.3	17.5
18.2	18.7	19.3	20.0	21.0	22.0	23.0	25.0	26.0	29.0

这 50 个数据中，最小值为 2.0，最大值为 29.0，全距＝29.0－2.0＝27.0。根据商品销售额的特点，将组距确定为 5，组数为 6。因为销售额是连续型变量，采用同限等距分



组，编制成组距式变量数列，同时可编制累计次数分布表及绘制次数分布图，分别见表3-13和如图3.4所示。

表3-13 50家超市销售额累计次数分布

商品销售额/百万元	组距	销售额分布		向上累计		向下累计	
		家数/个	比重	频数	频率	频数	频率
5以下	2.5	4	8%	4	8%	50	100%
5～10	7.5	10	20%	14	28%	46	92%
10～15	12.5	16	32%	30	60%	36	72%
15～20	17.5	13	26%	43	86%	20	40%
20～25	22.5	4	8%	47	94%	7	14%
25以上	27.5	3	6%	50	100%	3	6%
合计	—	50	100%	—	—	—	—

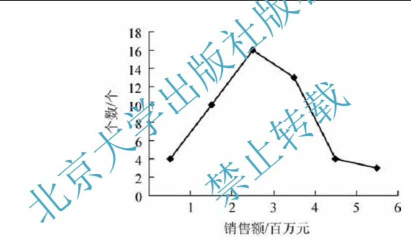


图3.4 50家超市销售额次数分布

整理后的资料可以清楚反映50家超市商品销售额分布特征。例如，销售额在10百万元以下的有28%，大于10百万元的有72%。销售额在10百万～15百万元的超市最多占32%。李华的超市去年销售收入990万元(9.9百万元)，尚处于下游但接近占比重最大的中游“主体”(更深入的特征描述须利用后面章节里介绍的统计工具)，考虑到自己仅创业3年，应该充分肯定成绩，同时看到差距，充满希望，继续努力。

4. 次数分布的主要类型

由于社会经济现象的复杂性，不同的统计总体有不同的次数分配，形成各种不同的次数分布类型。概括起来主要有3种类型：钟形分布、U形分布和J形分布。

(1) 钟形分布。钟形分布的特征是“两头小，中间大”，即靠近中间的变量值分布的次数多，靠近两边的变量值分布的次数少，如图3.5所示。社会经济现象中许多变量的分布属于钟形分布，如人的身高、体重、居民家庭人均收入、商品的价格等。

在统计学中，钟形分布根据是否对称又分为两种。一种是对称分布，也称为正态分布

(normal distribution)，其分布特征是中间变量值分布的次数最多，两侧变量值分布的次数随着与中间变量值距离的增大而渐次减少，两侧减少的速度相同，致使分布曲线呈对称分布，如图3.5(a)所示。客观实际中，许多社会现象统计总体的分布都趋于对称分布中的正态分布。正态分布是描述统计中的一种主要分布，它在社会经济统计分析中具有重要的意义。正态分布曲线下的面积和等于1或者100%，实际工作中，经常需要了解正态曲线下横轴上某一区间的面积占总面积的百分数。有3个区间的面积应用很多：一是区间($\mu-1\sigma$，$\mu+1\sigma$)的面积占总面积的68.27%，说明观察值(变量值)落在平均数±1标准差之内的概率为68.27%；二是区间($\mu-2\sigma$，$\mu+2\sigma$)的面积占总面积的95.45%，说明观察值(变量值)落在平均数±2标准差之内的概率为95.45%；三是区间($\mu-3\sigma$，$\mu+3\sigma$)的面积占总面积的99.73%，说明观察值(变量值)落在平均数±3标准差之内的概率为99.73%。

另一种是非对称分布，也称为偏态分布，其分布特征是中间变量值分布的次数最多，两侧变量值的分布次数逐渐减少，但是减少的速度快慢不同，致使分布曲线向一方偏斜，呈偏斜分布(skewed distribution)。根据分布曲线的偏斜方向将偏态分布分为两种：右偏态分布和左偏态分布，如图3.5(b)和图3.5(c)所示。

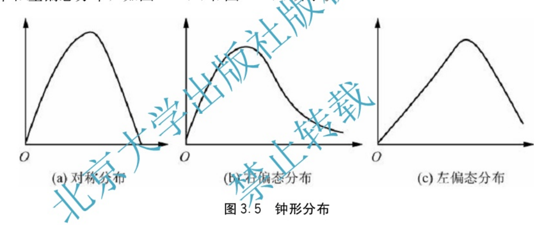


图3.5　钟形分布

(2) U形分布。U形分布的特征是“两头大，中间小”，靠近中间的变量值分布次数少，靠近两端的变量值分布次数多，如图3.6所示。例如，人口死亡率分布就是U形分布，幼儿和老人死亡率较高，而中青年死亡率较低；动物的死亡率也是U形分布。

(3) J形分布。J形分布的特征是“一头大，一头小”，大部分变量值集中在某一端点，如图3.7所示。J形分布有两种类型，即正J形分布和反J形分布。正J形分布是次数随着变量的增大而增多，如经济学中供给曲线和投资按利润率大小分布。反J形分布是次数随着变量增大而减少，如需求曲线。

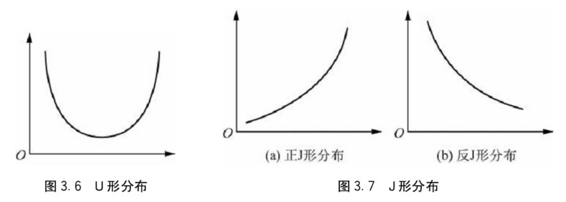


图3.6　U形分布

图3.7　J形分布



3.4 统计图表

统计表(statistical table)和统计图(statistical graph)是显示统计数据的重要工具。统计调查所获得的原始资料，经过统计整理，转化为系统化的、科学的统计资料，这些统计资料往往通过统计表和统计图表示出来。

3.4.1 统计表

统计表是统计工作中应用极其广泛的一种显示统计数据的工具。

1. 统计表的定义和作用

统计表是用来显示经过汇总加工后的综合统计资料的一种表格形式。在统计工作过程中，统计调查得到的原始资料，经过统计整理后得到可以说明社会经济现象的特征和规律的统计数据，将这些统计数据按一定的项目和顺序排列在表格上显示出来，就形成了统计表，如本章所编制的变量数列统计表。

采用统计表反映统计资料主要具有以下优点。

(1) 能使统计资料条理化、系统化，清晰地表达统计数据之间的相互联系。

(2) 能简洁、明了、紧凑地显示统计数据资料，具有叙述方式显示统计数据所无法比拟的优势。

(3) 便于计算和检查统计数据中数字的完整性和正确性。

2. 统计表的结构

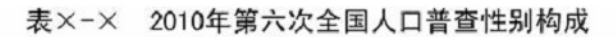
统计表的结构可以从内容和形式两方面进行分析。

(1) 从内容上看，统计表包括主词和宾词两个部分。主词就是统计表所要说明的总体及其组成部分。宾词就是统计表用来说明总体数量特征的各项统计指标及其数值。一般主词列在表的左方，列在横栏；宾词列在统计表的右方，列于纵栏。但是，统计表的主词和宾词排列的位置并不是固定不变的，有时考虑到资料显示的某些因素，也可以将主词与宾词的位置加以调整。

此外，统计表还可以包括补充资料、注解、资料来源、填表单位、填表人等附加内容。

(2) 从形式上看，统计表包括总标题、横行标题、纵栏标题和数字资料 4 个部分。总标题是统计表的名称，它简明扼要地概括了全表的基本内容，并指明时间和范围，一般放置于统计表格的上端正中。横行标题(也称横标目)是横行的名称，是指总体名称及其分组，一般放在表格的左方。纵栏标题(也称纵标目)是纵行的名称，是指用于说明总体及其分组的统计指标的名称，一般放在表格的上方。横行标题和纵栏标题共同说明填入表格中的统计数字所指的内容。数字资料填写在统计表格的核心部分，是表格中的指标数值，用来说明总体及其组成部分的数量特征，列在横行和纵栏的交叉处，如图 3.8 所示。

由图 3.8 可知，我国 2010 年人口是总体，横行标题中性别男和女是对总体的分组(即主词)，纵栏标题中人数与比重是反映总体规模和说明总体及各组数量特征的统计指标(即宾词)，表中的数字资料是各项指标的经济内容的具体表现。



总标题

性　别	人数/人	比　重
男	686 852 572	51.27%
女	652 872 280	48.73%
合计	1 339 724 852	100.00%

纵栏标题　横行标题　数字资料　主词　宾词

图 3.8　统计表结构

3. 统计表的分类

根据不同的分类标志，可以将统计表划分不同的类型。

（1）按用途不同，统计表可以分为调查表、整理表和分析表。调查表是指在统计调查中用于登记、搜集和表现原始统计资料的表格。整理表是指在统计整理过程中用于统计汇总和表现汇总结果的表格。分析表是指在统计分析中用于对汇总结果进行定量分析的表格。

（2）按总体分组的情况不同，统计表可以分为简单表、分组表和复合表。简单表是指主词未经任何分组形成的统计表，也称为一览表，通常是指仅列出总体各单位的名称或按时间先后顺序简单排列的统计表。简单表多用于统计整理的初级阶段，主要目的是搜集调查的原始资料，见表 3 - 14。

表 3 - 14　2007—2012 年我国手机网民数量情况

年　份	手机网民数量/万人	年　份	手机网民数量/万人
2007	5 040	2010	30 274
2008	11 760	2011	35 558
2009	23 344	2012	41 997

（资料来源：中国互联网络信息中心(CNNIC).）

分组表是指主词只按某一个标志进行分组形成的统计表。利用分组表可以揭示现象类型的特征、反映现象的内部结构及分析现象之间的相互关系，见表 3 - 15。

表 3 - 15　2010 年第六次全国人口普查我国人口数及其年龄构成情况

年　龄	年末人数/人	所占比重	年　龄	年末人数/人	所占比重
0～14 岁	222 459 737	16.60%	60 岁及以上	177 648 705	13.26%
15～59 岁	939 616 410	70.14%	合 计	1 339 724 852	100.00%

复合表是指主词按两个或两个以上的标志进行重叠式分组形成的统计表。利用复合表可以反映所研究的现象受几种因素的共同影响而发生的变化，可以更加准确地把握现象变



化的规律，详细地认识问题和说明问题，见表 3－16。分组标志增加时，组数就会成倍增加，而且分组太细也不利于研究现象的特征。可见，复合分组的标志并不是越多越好。

表 3－16　2010 年第六次全国人口普查全国按性别和受教育程度划分的人口情况

项　　目	人数/人	项　　目	人数/人
总计（6 岁及 6 岁以上）	111 601 269	大学专科	6 592 271
男	58 054 057	男	3 519 047
女	53 547 212	女	3 073 224
小 学	34 393 912	大学本科	4 470 446
男	16 287 856	男	2 453 738
女	18 106 056	女	2 016 708
初 中	48 412 954	研究生	427 805
男	25 955 085	男	242 051
女	22 457 869	女	185 754
高 中	17 303 881		
男	9 596 280		
女	7 707 601		

（资料来源：国家统计局.）

4. 统计表的设计

统计表一般按宾词进行设计。统计表的设计分为简单设计和复合设计。简单设计是将所要反映的指标平行排列，如图 3.9 所示。简单设计通过平行分组直接说明指标的内容。复合设计是将各个指标结合起来进行层叠排列，如图 3.10 所示。

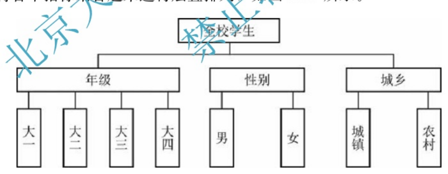


图 3.9　某高校学生平行分组体系

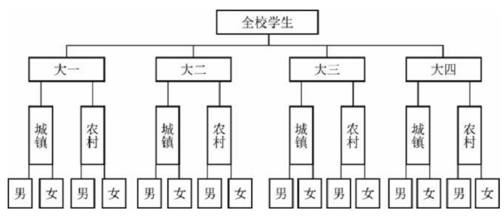


图 3.10　某高校学生复合分组体系

5. 统计表编制的规则

为了使统计表能更好地反映被研究对象的数量特征，便于比较分析，应遵循以下编制规则，使统计表更加科学、实用、简明、美观。

1）统计表表式设计规则

(1）表格样式。统计表应设计成由纵横交叉线条组成的长方形表格，长与宽之间保持适当的比例。统计表的上下两端的端线应当用粗线绘制，表中其他线条一律用细线绘制，表的左右两端习惯上均不画线，采用开口式。

(2）合计栏的设置。表中的横行“合计”一般列在最后一行(或最前一行)，表中纵栏的“合计”一般列在最前一栏。

(3）栏数的编号。如果栏数较多，应当按顺序编号，习惯上主词栏部分分别编以序号“甲、乙、丙、丁……”，宾词栏分别编以序号“(1)、(2)、(3)……”。若各栏中统计指标有一定的计算关系，还可以用算式表示之。

2）统计表内容设计规则

(1）标题设计。无论是总标题，还是横行、纵栏标题都应简明扼要，简练而又准确地表述统计资料的内容及所属的时间和空间范围。

(2）标题顺序。统计表中的主词与宾词的排列应尽量反映内容方面的逻辑关系。

(3）指标数值。表中数字应填写整齐，上下位置对齐。遇有相同数字应照写，不能用“同上”、“同左”等字样；当数字因小而忽略不计时，可填写为“0”；当缺某项数字资料时，可用符号“…”表示；不应有数字时，用符号“—”表示。

(4）计量单位。指标数字应有计算单位。如果表中的数字属同一计量单位，可将计量单位标在表的右上方；如果宾词的计量单位不同，可直接标注在指标名称的旁边或下方；如果主词的计量单位不同，可在横行标题后设计计量单位专栏。

(5）注解与资料来源。为保证统计资料的科学性与严肃性，在统计表下应注明资料来源，以便考察。必要时，在统计表下加注予以说明。

3.4.2 统计图

1. 统计图的作用和结构

统计表的资料用几何图形或图案等形式表示即成为统计图。统计图也是显示统计数据的一种形式。如果说利用统计表能够集中有序地表现统计资料，那么利用统计图则能够更加形象具体、简明生动、一目了然地展示统计资料，便于人们直观地认识事物的特征。统计图可以表示现象之间的对比关系，揭示总体结构及其变化发展趋势，分析现象之间的依存关系等。随着计算机技术不断发展，计算机制图功能日益强大，使得统计图的制作更加方便和精确。

统计图由标题、图域、标目、尺度和图例 5 部分构成。标题，每个图都应有标题，标题要简明确切，通常包括内容、时间和地点，其位置在图域之外，一般放在图域的下面。图域，图域的长宽之比一般以 7∶5 为美观，圆形图除外。标目，纵横两轴应有标目，即



纵标目和横标目，并注明度量衡单位。尺度，纵横两轴都有尺度，横轴尺度从左至右，纵轴尺度自下而上，数值一律由小而大，尺度间隔要宽松，用算术尺度时，等长的距离应代表相等的数量。图例，用不同线条或颜色代表不同事物时，需用图例说明。

2. 常见的统计图的绘制

统计图的形式多种多样，本节主要介绍以下常用的 3 种图形。

（1）条形图(bar chart)。条形图也称为柱形图，是用宽度相同的条形的高度或长度表示统计数据大小或多少的一种图形。条形图可以横置也可以纵置，纵置时又称为柱形图，即当各类别放在纵轴时，称为条形图；当各类别放在横轴时，称为柱形图。条形图主要用于说明或比较同一指标在不同时间、地点、单位的变化发展情况。例如，表 3 - 17 与图 3.11 反映了我国 2008—2012 年普通高等教育招生人数的变动情况。

表 3 - 17　我国 2008—2012 年普通高等教育招生人数



年　份	我国普通高等教育招生人数/万人	年　份	我国普通高等教育招生人数/万人
2008	607.66	2011	681.50
2009	639.49	2012	688.83
2010	661.76		

（资料来源：中华人民共和国教育部网站．）

图 3.11　我国 2008—2012 年普通高等教育招生人数条形图

（2）圆形图(circle chart)。圆形图又称为饼图(pie graph)，是用圆形和圆内扇形的面积大小表示统计指标数值大小的一种图形。它用于表示总体中各组成部分所占的比例，揭示现象的内部结构及其变化。在绘制圆形图时，总体中各部分所占的百分比用圆内的各个扇形面积表示，这些扇形的中心角度是按各部分百分比占 360°的相应比例确定的。例如，表 3 - 18 和图 3.12 显示了 2012 年我国农业人口就业结构状况。

从图 3.12 中可以看出，对农民来说，非农就业已经成为主流方向，超过了在农业领域的就业数量。

表 3-18　2012 年我国农业人口就业结构

农业人口就业流向	所占比重	农业人口就业流向	所占比重
完全从事非农工作	46.6%	兼务农业和非农职业	13.4%
完全从事农业劳动	40%	合计	100%

（资料来源：中国社会科学院社会学研究所、社会科学文献出版社《2012 年社会蓝皮书》.）

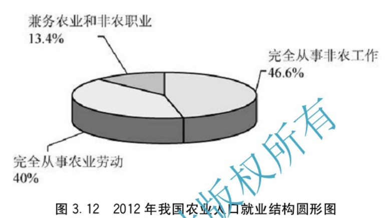


图 3.12　2012 年我国农业人口就业结构圆形图

（3）曲线图(curvilinear chart)。曲线图又称为折线图，是利用曲线的升、降变化表示统计指标数值变化的一种图形。它用于分析社会经济现象发展变化的规律、趋势及现象之间的依存关系。

绘制曲线图时，如果是某一现象随时间变化的显示，则应将时间绘制在横坐标轴上，指标绘制在纵坐标轴上；如果是两个现象依存关系的显示，则一般将表示原因的指标绘制在横坐标轴上，表示结果的指标绘制在纵坐标轴上。

例如，2007—2012 年我国城乡居民家庭人均收入见表 3-19 和如图 3.13 所示。

表 3-19　2007—2012 年我国城乡居民家庭人均收入情况

年　　份	农村居民/元	城镇居民/元	年　　份	农村居民/元	城镇居民/元
2007	4 140	13 786	2010	5 919	19 109
2008	4 761	15 781	2011	6 977	21 810
2009	5 153	17 175	2012	7 917	24 565

（资料来源：中华人民共和国 2012 年国民经济和社会发展统计公报.）

从图 3.13 可以清楚地看到，城乡居民的家庭人均收入逐年提高，城镇居民的家庭人均收入高于农村，而且这种差距有扩大的趋势。

绘制曲线图时应注意：图形的长宽比例要适当，一般为横轴略大于纵轴，其长宽比例大致为 10∶7，图形过扁或过于瘦高，不仅不美观，而且会给人造成视觉上的错觉，不便于对数据变化的理解。

3. 绘制统计图的基本要求和原则

为了使统计图能准确、生动地反映被研究对象的数量特征，在编制统计图时应注意以



下编制规则。

(1) 各种图形的适用条件不同，应选择恰当的图形来表示统计数据。

(2) 图的标题应简明扼要地说明所要表达的内容。图的标题位于表的下方，字体是最大的。

(3) 有纵横轴的图形，横轴表示研究对象，尺度要等距，自左至右，由小到大。纵轴一般表示现象出现的频数或频率，从零开始等距分点，由下至上，从小到大。数字位于左侧，要注明单位。

(4) 线条粗细有差别，图形线条最粗，坐标线条较细。

(5) 在同一图形上比较多个事物时取的尺度要相同，比较的对象不宜过多。

(6) 图形上尽量不要写数字，如要说明则应该写在图注中。

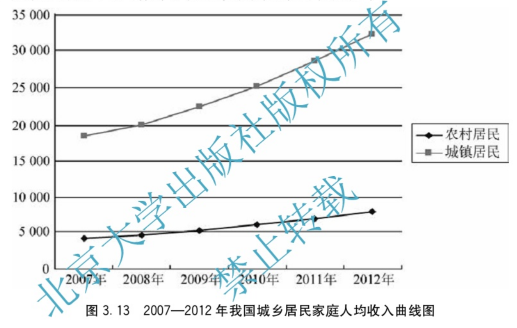


图 3.13 2007—2012 年我国城乡居民家庭人均收入曲线图

本章小结

本章共有 4 节内容，按照统计整理的主要步骤进行了阐述。

统计整理是一个承上启下的工作过程。它根据统计研究的目的，对统计调查所取得的资料进行科学加工，使之成为系统化、条理化，能反映现象总体特征的综合资料。统计整理的步骤主要包括审核原始资料、进行分组汇总和编制统计图表。统计资料的审核主要包括对资料准确性、及时性和完整性的审核。

统计分组是统计整理的关键，其主要作用为 3 个方面。统计分组从不同角度可划分为不同的类型。统计分组的关键在于正确选择分组标志和划分各组界限。统计分组方法主要有品质标志分组和变量标志分组两种。

在统计分组的基础上，将总体的所有单位按组进行归类整理并按一定顺序排列，计算出各组的单位数，形成一个反映总体中各单位在各组中的分布情况的数列，称为分配数列。分配数列有品质数列和变量数列两种。而变量数列又有单项式和组距式两类。其中，组距式变量数列的编制方法是本章的重点及难点。

为了统计工作的需要，也为了形象、直观地显示统计数据，需要编制统计表和绘制统计图。统计表的结构从内容上包括主词和宾词两个部分；从形式上包括总标题、横行标题、纵栏标题和数字资料4个部分。根据不同的分类标志，可以分为简单表、分组表和复合表。常用的统计图有条形图、圆形图和曲线图3种。

应用与拓展

2013年中国网民结构特征

从1998年起，中国互联网络信息中心于每年1月和7月发布《中国互联网络发展状况统计报告》。在第33次（2014年1月）调查中，通过对抽样样本的调研，汇总出2013年中国网民具有以下结构特征。

1. 性别

截至2013年12月，中国网民男女比例为56∶44，与2012年情况基本保持一致。庞大的网民基数影响下中国网民性别比例保持基本稳定，如图3.14所示。

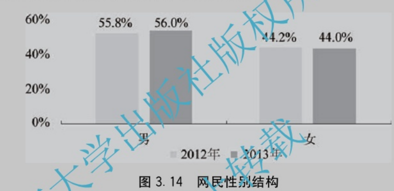


图3.14　网民性别结构

2. 年龄

如图3.15所示，截至2013年12月，我国20～29岁年龄段网民的比例为31.2%，在整体网民中占比最大，和2012年底网民结构一致。而低龄和高龄网民略有提升，这意味着互联网的普及继续深入。

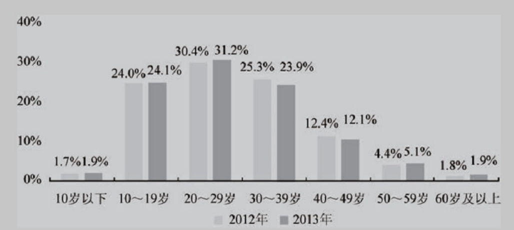
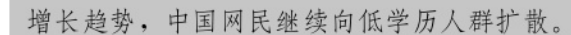


图3.15　网民年龄结构

3. 学历

如图3.16所示，截至2013年12月，高中及以上学历人群中互联网普及率已经到较高水平，未来进一步增长空间有限。2013年，小学及以下学历人群的占比为11.9%，相比2012年有所上升，保持





增长趋势，中国网民继续向低学历人群扩散。



图 3.16 网民学历结构

4. 网民职业结构

如图 3.17 所示，学生依然是是中国网民中最大的群体，占比 25.5%，互联网普及率在该群体中已经处于高位；个体户/自由职业者构成网民第二大群体，占比 18.6%；企业公司中管理人员占比为 2.5%，一般职员占比为 11.4%。

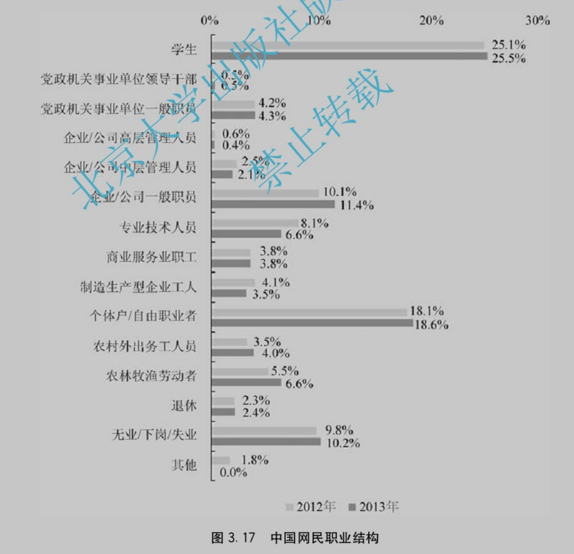


图 3.17 中国网民职业结构

5. 月收入结构

如图 3.18 所示，月收入为 2 001～3 000 元和 3 001～5 000 元的上网群体规模最大，在总体网民中占比分别为 17.8%和 15.8%；500 元以下及无收入人群占比为 20.8%。

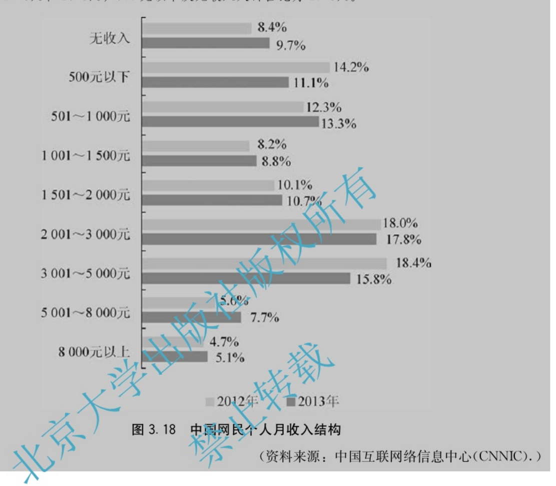


图 3.18　中国网民个人月收入结构

（资料来源：中国互联网络信息中心(CNNIC).）



习题与实训

一、单项选择题

1. 对资料进行审核，主要是审核资料的(　　)。

A. 准确性、及时性和完整性　　B. 全面性、及时性和完整性

C. 准确性、及时性和经济性　　D. 全面性、时效性和经济性

2. 某连续变量数列，其首组为 100 以下，若其相邻的组中值为 130，则首组的组中值为(　　)。

A. 60　　B. 70　　C. 80　　D. 100

3. 统计分组的目的是体现(　　)。

A. 组内同质性、组间差异性　　B. 组内差异性、组间同质性

C. 组内同质性、组间同质性　　D. 组内差异性、组间差异性



4. 次数分配数列是(　　)。

A. 按数量标志分组形成的数列

B. 按品质标志分组形成的数列

C. 按统计指标分组形成的数列

D. 按数量标志和品质标志分组形成的数列

5. 连续变量分组，若第一组为100以下，第二组为100～150，第三组为150～200，第四组为200以上，则数据(　　)。

A. 100在第一组　　B. 150在第二组

C. 200在第三组　　D. 200在第四组

6. 100名员工的日产量资料显示，最高日产量为38件，最低日产量为19件。若要对这100名员工按日产量进行等距分组，组数为5，则组距应为(　　)。

A. 3件　　B. 4件　　C. 8件　　D. 10件

7. 某班级学生成绩最低分数是51分，最高分数是100分，若分成5组，则下面较优者是(　　)。

A. 第一组为“60以下”，最后一组为“90以上”

B. 第一组为“50～60”，最后一组为“90以上”

C. 第一组为“60以下”，最后一组为“90～100”

D. 第一组为“50～60”，最后一组为“90～100”

8. 某公司对其所属企业的生产计划完成百分比采用如下分组，正确的是(　　)。

A. 80%～89%，90%～99%，100%～109%，110%以上

B. 80%以下，80.1%～90%，90.1%～100%，100.1%～110%

C. 90%以下，90%～100%，100%～110%，110%以上

D. 85%以下，85%～95%，95%～105%，105%～115%

9. 下面关于“组限”的描述不正确的是(　　)。

A. 特殊的界限点必须作为组限　　B. 连续性变量相邻组的组限必须重叠

C. 最小组的下限要略低于最小变量值　　D. 最大组的上限要略小于最大变量值

10. 将我国国内生产总值按地区划分为东部、中部、西部、东北4部分形成的统计表是(　　)。

A. 简单表　　B. 分组表　　C. 复合表　　D. 调查表

二、多项选择题

1. 下列按数量标志分组的有(　　)。

A. 学生按性别分组　　B. 家庭按收入水平分组

C. 工人按日装配零件数分组　　D. 企业按出口额分组

E. 员工按销售数量分组

2. 下列可以编制单项式数列的有(　　)。

A. 家庭按拥有的汽车数分组

B. 居民按人均可支配收入分组

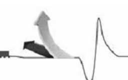
C. 出口企业按与之有贸易联系的国家个数分组

D. 学生按身高分组
E. 教师按职称分组
3. 在组距数列中，组中值(　　)。
A. 是上限和下限之间的中点数值
B. 用来代表各组标志值的平均水平
C. 在开放式分组中无法确定
D. 在开放式分组中，可以参照相邻组的组距来确定
E. 就是组平均数
4. 在次数分配数列中(　　)。
A. 总次数一定，频数和频率成反比
B. 各组的频数之和等于100
C. 各组频率大于0，频率之和等于1
D. 频数越小，则该组的标志值所起的作用越小
E. 频率又称为次数
5. 表3-20中的数列所属类型为(　　)。

表3-20　学生成绩分布情况

成绩/分	学生数/人	成绩/分	学生数/人
60以下	2	80～90	10
60～70	12	90以上	5
70～80	21	合 计	50

A. 品质分配数列　　B. 变量分配数列
C. 组距变量分配数列　　D. 等距变量分配数列
E. 次数分配数列
6. 正确选择分组标志的原则是(　　)。
A. 要根据事物发展的规律选择分组标志
B. 选择最能体现事物本质特征的标志作为分组标志
C. 要根据研究的目的和任务选择分组标志
D. 根据数量标志和品质标志的不同选择分组标志
E. 要结合现象的历史条件和经济条件选择分组标志
7. 下列是变量分配数列的有(　　)。
A. 大学生按年龄的分配　　B. 超市按销售额的分配
C. 出口企业按出口目的地的分配　　D. 居民按消费支出水平的分配
E. 工人按生产每一零件时间消耗的分配
8. 从内容看，所有的统计表都是由(　　)组成的。
A. 横行标题　　B. 纵栏标题　　C. 宾词
D. 主词　　E. 数字资料



9. 下列只能编制组距数列的有(　　)。

A. 家庭按拥有微型计算机数分组　　B. 职工按月工资额分组

C. 商场按营业收入分组　　D. 学生按每周上网时间分组

E. 城市按年地区生产总值分组

10. 某一分配数列见表3-21。

表3-21　员工每天拼装玩具情况

实习员工每天拼装玩具数/个	员工人数/人	实习员工每天拼装玩具数/个	员工人数/人
5以下	7	15～19	20
5～9	18	20以上	3
10～14	52	合计	100

以下说法正确的是(　　)。

A. 第四组的组中值是17　　B. 最后一组为开口组

C. 第三组的组距是5　　D. 第一组的下限是5

E. 第五组的频率是30%

三、判断题

1. 统计分组的关键是确定组距和组数。(　　)

2. 离散型变量只适合于单项式分组。(　　)

3. 凡是将总体按标志值分组所形成的数列，都叫变量数列。(　　)

4. 次数密度就是各组单位组距内分布的次数。(　　)

5. 连续性变量只能编制组距式变量数列；离散型变量既可以编制单项式变量数列，也可以编制组距式变量数列。(　　)

6. 统计表按总体分组的情况不同可分为调查表、整理表和分析表。(　　)

7. 变量数列中的开口组不能确定组中值。(　　)

8. 累计次数的特点：同一数值的向上累计和向下累计次数之和等于1。(　　)

9. U形分布的特征是“两头大，中间小”。(　　)

10. 在组距分组时，对于连续型变量，相邻两组的组限必须是重叠的。(　　)

四、填空题

1. 统计整理的内容主要包括拟定统计整理纲要、________、________、汇总、编制统计图表和进行统计资料的保管与积累。

2. 在变量为离散变量并且变量值的变动范围不大的情况下，适合编制________分配数列。

3. 分配数列按分组标志特征的不同，分为________和________。

4. 在组距数列中，表示各组界限的变量值称为________，下限是指________的变量值，上限是指________的变量值。各组上限与下限之间的中点数值称为________。

5. 组距与组数关系密切，同一现象的变量分组，组距与组数成________。

6. 连续变量分组，相邻两组的上限与下限通常是相同的，为了避免计算时总体单位数出现错误，统计各组单位数时遵循这一原则：将到达上限值的单位数计入下一组内。这一原则被称为________。

7. 组距式分组根据其分组的组距是否相等可分为________分组和________分组。

8. 次数分配是由________和________两个要素构成的。表示各组单位数的次数又称为________，各组次数与总次数之比称为________。

9. 缺下限的开口组组中值计算公式为________。

10. 主词只按某一个标志进行分组形成的统计表称为________表。

五、应用能力训练题

1. 以下是某企业30名工人在一个工作日内生产的零件数量。请根据给出的资料编制单项式变量数列。

20	21	21	24	23	22	20	21	22	23
24	20	24	21	22	22	23	24	22	21
22	23	22	21	21	22	22	23	22	23

2. 对某企业30名工人完成劳动定额(%)的情况进行调查，某原始资料如下。

98	81	95	84	93	86	91	102	100	103
105	100	104	108	107	108	106	109	112	114
109	117	125	115	120	119	118	116	129	113

(1) 请根据给出的资料编制组距变量数列。

(2) 编制向上累计频数(频率)和向下累计频数(频率)分布。

3. 某企业工人日产量资料见表3-22。

表3-22 某企业工人日产量

按日产量分组/件	工人数	按日产量分组/件	工人数
50～60	6	90～100	15
60～70	9	100～110	18
70～80	12	110～120	20
80～90	14	120～130	6

要求：

(1) 指出上述变量数列属于哪一种变量数列，这个数列说明什么问题?

(2) 指出统计表中的变量、变量值、上限、下限、次数、频率、总体单位总数。

(3) 计算组距和各组组中值。

4. 某班学生统计学考试成绩(单位：分)如下。

93	50	78	85	66	71	63	83	52	95
78	72	85	78	82	90	80	55	95	67
72	85	77	70	90	70	76	69	58	89
80	61	67	99	89	63	78	74	82	88
98	62	81	44	76	86	73	83	85	81



根据上述资料，要求：

（1）编制组距数列，说明每一组的上下限、组中值。

（2）绘制次数分布曲线图，据此分析成绩分布的特点。

（3）编制累计频数分布表，并计算60分以下及80分以上的人数。

5. 某市30家企业的固定资产数据(单位：万元)如下。

285　340　386　417　485　615　562　620　622　648
655　690　721　743　795　815　840　878　925　930
955　1 140　1 200　1 225　1 240　1 342　1 331　1 421　1 540　1 624

要求：分别编制组距为200和300的等距数列，并分析哪种分组方法较好?

6. 某地区各工业企业工人数(单位：人)如下。

322　674　339　357　346　295　465　355
332　316　453　442　417　587　369　545
323　430　560　528　333　311　410　604
281　461　432　421　484　473　392　354

根据上面的数据编制组距为100的分布数列，并绘制次数分布图。

7. 某地区200家企业的有关资料如下。

2 000人以下的企业中全民企业10家，集体企业5家，合资企业3家；2 000～3 000人的企业中全民企业20家，集体企业34家，合资企业4家；3 000～4 000人的企业中全民企业15家，集体企业15家，合资企业2家；4 000～5 000人的企业中全民企业20家，集体企业15家，合资企业1家；5 000～6 000人的企业中全民企业20家，集体企业10家，合资企业1家；6 000人以上的企业中全民企业15家，集体企业10家，合资企业0家。

根据所给资料编制统计表，要求：

（1）按品质标志分组编制简单分组表；

（2）按变量标志分组编制变量数列。

（3）设计复合表。

8. 为了确定灯泡的使用寿命(单位：小时)，在一批灯泡中随机抽取100只进行测试，所得结果如下。

700　716　728　719　685　709　691　684　705　718
706　715　712　722　691　708　690　692　707　701
708　729　694　681　695　685　706　661　735　665
668　710　693　697　674　658　698　666　696　698
706　692　691　747　699　682　698　700　710　722
694　690　736　689　696　651　673　749　708　727
688　689　683　685　702　741　698　713　676　702
701　671　718　707　683　717　733　712　683　692
693　697　664　681　721　720　677　679　695　691
713　699　725　726　704　729　703　696　717　688

根据上述资料，要求：

(1) 整理成组距数列(以组距为10进行等距分组)，编制分布表。

(2) 绘制直方图和次数分配曲线。

9. 根据对所在大学的学生进行上网调查所获得的原始资料进行统计整理，设计各种统计表并绘制统计图以反映学生上网时间、上网目的、上网途径等的分布情况，在此基础上了解学生上网的综合情况，并编写统计调查分析报告。

10. 利用第2章中对所在的班级总体进行统计调查所搜集的资料，要求：

(1) 按性别分组，编制品质分配数列。

(2) 按年龄分组，编制单项式变量数列。

(3) 按身高分组，编制组距式变量数列。



第4章 综合指标

理论目标

(1) 理解总量指标、相对指标和平均指标的意义、种类和作用。
(2) 掌握相对指标的计算方法和应用特点。
(3) 掌握平均指标的计算方法和适用范围。
(4) 理解标志变异指标的意义、作用和计算。

能力目标

(1) 能采用对比分析法(利用相对指标)对客观现象进行描述和评价。
(2) 能根据客观实际正确选择平均指标反映现象集中趋势。
(3) 能够运用标志变异指标进行平均数代表性的对比分析。

做一做

BLOOMY 公司是拥有 12 家企业的生产公司，该公司“十二五”计划总产量 400 万台，2010 年产量 80 万台，2011 年产量 85 万台，而据资料，同行业的竞争对手 PROPER 公司年产量可达 102 万台。2012 年，公司加大改革措施，进行战略调整，当年即取得显效。2012 年具体生产情况见表 4-1。

据此资料可以判断，2012 年公司取得了辉煌业绩，作为公司管理人员，在年终总结报告中，应该怎样生动、丰富又有说服力地表达生产业绩的“好”?

表 4-1 BLOOMY 公司 2012 年下属企业生产情况

按产量计划完成程度分组	企业数/家	计划产量/万台
100%～110%	2	13
110%～120%	7	53
120%～130%	3	24
—	12	90

■ 想一想

表 4-1 中给出的资料是 BLOOMY 公司 2012 年下属 12 家企业的生产情况，应如何根据此表描述公司 2012 年总的产量计划完成情况？为表达公司显著业绩，应该全方位、多角度、公正、科学地阐述。根据给出的资料，如何用 BLOOMY 公司 2012 年实际总产量正确地与计划产量相比、与竞争对手比较、与过去几年比较、与“十二五”计划相比较以达到我们的目标？本章将介绍一种应用极其广泛的描述现象特征的统计方法——综合指标法，包括总量指标、相对指标和平均指标分析方法。

4.1 总量指标

4.1.1 总量指标的意义和种类

1. 总量指标的意义

总量指标是以绝对数形式反映一定客观现象总体在具体时间和地点条件下的总规模、总水平或工作总量的统计指标，又称统计绝对数(absolute number)。例如，全国人口总数可以反映全国人口总规模，某市某年国内生产总值反映该城市全年 3 次产业增加值总量。总量指标也就是前面所述的数量指标，总量指标是最基本的综合指标。

在社会经济统计中，总量指标具有重要作用。它可以通过人口总数、劳动力数量、国内生产总值、各种主要产品产量等总量指标，反映一个国家的基本国情和国力；可以为我们编制长期规划、进行监督管理提供依据；总量指标是综合指标的基本形式，它是形成相对指标和平均指标的必备基础。因为，相对指标和平均指标都是总量指标的派生指标。因此，总量指标的计算正确与否直接影响到其他指标计算的正确性。

2. 总量指标的种类

(1) 按反映的内容不同，总量指标可分为总体单位总量和总体标志总量。总体单位总量表示的是一个总体内所包含的总体单位总数，即总体本身规模的大小。总体标志总量是总体各单位某种数量标志值的总和，是说明总体特征的总数量。例如，对某城市居民的粮食消费情况进行研究时，该城市居住的人口总数便是总体单位总量指标，居民的粮食消费总数则是总体标志总量指标。



总体单位总量和总体标志总量并不是一成不变的，随着研究目的和研究对象的不同而变化。例如，当以全国为研究总体来观察某城市的人口数时，上述“人口数”指标就是总体标志总量指标；当研究目的为某市的粮食销售价格时，“粮食消费总数”便是总体单位总量指标。

（2）按反映的时间状况不同，总量指标可分为时期指标和时点指标。时期指标是反映客观现象总体在一段时期内活动过程的总数量，即流量指标。例如，产品产量、原材料消耗量、一定时期的基本建设投资额、商品流转额及出生人口数等。这一指标说明时期现象在相应时期内发生的总量是多少。这种现象是连续不断地发生的，每个时期的累计数就表明现象在该时期整个活动过程的总成果。

时点指标是反映客观现象总体在某一时点(瞬间)上状况的总量指标，即存量指标。例如，期初或期末固定资产额、原材料库存量及人口总数等。这一指标说明时点现象在相应时点上所达到的总量水平的大小。

时期指标和时点指标都是总量指标，这是它们的共同点，但它们又有明显的区别。时期指标数值的大小与时间长短有直接关系，各时期数值可以相加，资料搜集通过经常性调查取得。例如，1 个月的产量肯定比 1 年的产量少，而 1 年的总产量，是由 1 年中每天的产量连续登记汇总得到的。时点指标是说明现象在某一时刻状态上某种标志值的总量，其大小与时间长短无直接关系，各指标数值相加无意义，资料搜集通过一次性调查取得。例如，某班级学生第一天出勤 50 人，第二天出勤 49 人，不能把两天出勤人数相加得出出勤人数 99 人。出勤人数多少的变化与时间的长短无关。

4.1.2 总量指标的计量单位

总量指标是社会经济现象总量的具体表现，采用绝对数形式，必须有一定的计量单位。计量的准确与否直接影响总量指标反映客观现象的准确程度。根据事物的性质特点和研究任务的不同，总量指标的计量单位可分为以下 3 种类型。

1. 实物单位

实物单位是根据事物的自然属性和特点所采用的计量单位。例如，人口总数以“人”进行计量，钢铁产量以“千克”进行计量，等等。它可以明显而又准确地反映事物的使用价值量的大小，但缺乏综合性。实物单位可进一步分为以下几种。

（1）自然单位。它是根据被研究现象的自然状况来统计其数量的一种计量单位。例如，职工人数以“人”、汽车以“辆”、鞋以“双”为单位，等等。

（2）度量衡单位。它是按照统一的度量衡制度的规定来度量客观事物数量的一种计量单位。一般采用国际公制度量衡单位进行计量。例如，煤炭产量以“千克”、布产量以“米”、天然气产量以“立方米”为计量单位，等等。采用度量衡单位的原因是有些现象无法采用自然计量单位来表明其数量，如钢铁、粮食等。另外，有些现象虽然可以采用自然计量单位，但不如度量衡单位准确。统一度量衡制度是准确反映客观事物数量的前提。

（3）双重和复合单位。它是综合两种或两种以上实物单位来完全反映事物实际数量或

实际能力的计量单位。双重单位常常以相除的方法将两个单位结合在一起。例如，电动机功率以“千瓦/台”为计量单位。复合单位则是用相乘的方法将两种计量单位有机地结合在一起来表示事物的数量。例如，国内旅游出游人数以“人次”为计算单位，旅客运输周转量以“人·千米”为计量单位，等等。

(4) 标准实物单位。它是对名称和用途基本相同，但效能不同的事物，按照其物理、化学或其他性质折算成标准实物的计量单位。例如，将发热量不同的煤折合为每千克发热量为29.307 6 kJ的标准煤计算产量；各种不同含量的化肥用折纯法折算为100%含量的化肥进行计算；集装箱吞吐量按20英尺货柜(twenty-foot equivalent unit，TEU)为标准箱进行折算，等等。

用实物单位计算的总量指标，称为实物指标。实物指标的优点是能直接反映商品的使用价值和现象的具体内容，因而能直接表明事物的规模水平。其缺点是综合性能差，性质不同的计量单位无法进行汇总。

2. 价值单位

价值单位是采用货币作为价值尺度来度量劳动成果或社会财富的一种计量单位，如工资总额、国内生产总值、社会商品零售额、产品进出口总额等。

用价值单位计算的总量指标，称为价值指标。价值指标的优点是使不能直接相加的产品总量或商品数量过渡到能够加总，可以综合说明具有不同使用价值的产品总量或商品销售量等的规模或总水平。因此，价值指标具有广泛的综合性和概括能力。其缺点是价值指标脱离了物质内容，比较抽象。

价值指标按计算价格的不同分为两种。一是按现行价格计算的价值指标。例如，工业总产值是用报告期内销售产品的实际出厂价格计算的，反映现象实际的水平，是研究国民经济现实经济关系和一些重要比例的依据。二是按不变价格计算的价值指标，它消除了价格变动因素的影响，可以真实地反映事物发展的水平和规模。

3. 劳动量单位

劳动量单位是采用劳动时间来度量劳动成果的一种计量单位，如出勤工日、定额工时、实际工日等。劳动量指标是评价劳动时间利用程度和计算劳动生产率的依据，同时也是企业编制生产计划和检查生产计划的依据。

4.1.3 总量指标的计算和应用

1. 总量指标的计算

总量指标的计算方法有直接计量法、推算法和估算法。

直接计量法是指对研究对象进行直接记数、测量后将总量指标的数值计算出来的方法。这种方法要求对总体的所有单位都进行登记，并汇总出所需要的资料，如商品流转统计报表中的库存量、人口普查资料等。

在总量指标不能直接计算或不必要直接计算的情况下，可采用推算法和估算法。常用



的推算法包括以下几种。

（1）因素关系推算法。即利用社会经济现象的各个影响因素之间的关系，根据已知因素来推算未知因素的方法。例如，“销售额＝单价×销售量”，这一关系式中的某两项已知就可推算另一未知因素的数值。

（2）比例关系推算法。即利用各种相关资料的比例关系进行推算的一种方法。例如，某超市零售总额中，团体消费品零售额历年均占零售总额的10%，已知2012年该超市零售总额为5 000万元，利用比例关系推算法，则该超市2012年团体消费品零售额为500万元(5 000万元×10%)。

（3）平衡关系推算法。即利用各种平衡关系来推算未知指标的方法。例如，“期初库存＋本期购进＝本期销售＋期末库存”，这一关系式中的3项已知就可以利用平衡关系推算另一未知因素的数值。

估算法是指运用抽样推断的方法估算总量指标，具体内容将在后面章节中介绍。

2. 总量指标的应用

在实际统计工作中，经常运用统计调查的资料汇总计算总量指标，此时，应注意以下问题。

（1）科学确定总量指标的含义、范围和计算方法。总量指标的计算并非单纯的汇总技术问题。例如，计算工业企业增加值时，首先应对“工业企业”含义加以确切界定，然后是增加值包括的范围和计算方法问题，确定后才能进行正确的统计。

（2）计算实物总量指标时，要注意现象的同类性。实物指标的同类性是由产品的使用价值决定的。只有使用价值相同的产品才能汇总计算总量指标。把不同使用价值的产品产量简单加总是没有意义的。例如，简单地把煤炭、石油、粮食等产品产量相加就没有确切的经济含义。但是对现象同类性的要求不能绝对化。例如，在计算粮食产量时，就不必考虑粮食是小麦、水稻，还是玉米、高粱。再如，计算货物运输量总量时，产品的同类性就不成为计算的条件，因为它只要求通过货物的重量和里程，计算货物量和货物周转量。

（3）必须使用统一计量单位。总量指标的计量单位，必须按照国家统一的计量单位进行计量。例如，在国际贸易、市场营销等经济管理工作中，为了准确地反映客观事物的总量，在进行各个时期、各个地区甚至国际间资料的对比时，应规定各种统一标准计量单位和统一度量衡制度。只有这样，才不会造成统计计量方面的差错或混乱，才能客观地统计社会经济现象总体的数量。

我国从1991年起统一使用国际单位制为基础的法定计量单位制度，促进了实物总量指标的准确统计。

（4）总量指标在实际应用时，一般不单独使用。一是可以与同类总量指标相加减。例如，本月产量100万台，比去年同期增长了16万台，截至本月累计生产380万台。二是与相对指标和平均指标结合应用。例如，本月产量100万台，比上月增长了20%，工人人均月产量达到300台。

4.2 相对指标

4.2.1 相对指标的意义和种类

总量指标虽然可以综合反映社会经济现象的规模、水平和工作总量，但由于现象总体的复杂性，仅根据总量指标难以对客观事物做出正确的判断。相对指标是在总量指标的基础上进行对比而产生的统计分析指标，它有利于反映现象之间的联系状况。

1. 相对指标的意义

相对指标又称统计相对数(relative number)，它是指两个相互联系的统计指标之比。例如，人口的性别比例、人口的出生率和死亡率、资金利润率、经济发展速度、人均国内生产总值、失业率、物价指数等，都是相对指标。相对指标的特点在于指标数值的抽象化。就是说，它是把两个具体数值加以概括化、所形成的抽象化的数值。通过这种抽象化，使人们对客观现象的数量特征有一个更清晰的认识。

在社会经济生活、国民经济管理和统计研究中，相对指标应用很广泛，具有十分重要的意义。首先，相对指标能够反映社会经济现象之间的数量对比关系。因为，发展规模的大小、计划执行的好坏、变化速度的快慢、各种比例协调与否，只有通过比较才能反映出来。其次，相对指标把社会经济现象的绝对差异抽象化，使原来不能直接对比的指标可以进行对比。例如，考察不同地区居民生活的富裕程度时，由于各地区客观条件不同，不能用总量指标直接对比。但如果都以各自的食品支出总额和消费支出总额指标作为依据，计算结构相对指标恩格尔系数，就可以进行比较和深入分析。

2. 相对指标的表现形式

相对指标的表现形式有两种：一种是无名数，另一种是有名数。无名数是一种抽象化的数值，多用倍数或系数、成数、百分数和千分数表示。

(1) 倍数和系数。它们是将对比基数抽象化为1来计算的相对数。当分子数值比分母数值大很多时，常用倍数表示。当分子数值和分母数值差别不大时，常用系数表示，系数可以大于1，也可以小于1。

(2) 成数。它是将对比基数抽象化为10来计算的相对数。例如，粮食产量增加两成，即增长2/10。

(3) 百分数和千分数。百分数(%)是将对比基数抽象化为100来计算的相对数，是相对指标中最常用的表现形式。百分点是百分数的另一种表述形式，它以1%为单位，即1个百分点等于1%。它在两个百分数相减的情况下应用。例如，远东银行活期储蓄利率为2.1%，现在下调一个百分点，这说明现在银行活期储蓄利率为1.1%。

千分数(‰)是将对比基数抽象化为1 000而计算的相对数。一般在两个数值对比时，当分子比分母的数值小很多时，则用千分数表示。例如，我国2012年人口出生率为12.10‰；死亡人口966万人，死亡率为7.15‰；自然增长率为4.95‰。



（4）翻番数。翻番数是指两个相比较的数值中，一个数是另一个数的“2^n”倍，则 n 是番数。例如，某地区 2012 年的工业总产值为 240 亿元，计划 2013 年翻一番，则该地区 2013 年工业总产值应达到 480 亿元；计划翻两番，则该地区 2013 年工业总产值应达到 960 亿元；翻三番则应达到 1 920 亿元。

有名数主要用于计算部分强度相对指标。例如，人口密度用“人/平方千米”，居民人均生活费收入用“元/人”等表示。通常是将进行对比的两个经济指标的计量单位同时使用，以双重单位表示。

3. 相对指标的种类

相对指标既然是两个有联系的指标之比，那么，哪两种指标之间有联系？把这两种指标进行对比之后能说明什么问题？根据这个标准，有联系的指标主要有下列几种：结构相对指标、比例相对指标、比较相对指标、动态相对指标、强度相对指标和计划完成程度相对指标。这些指标从不同方面表明现象的相对水平，表明现象的结构组成、发展程度、比例关系等。

4.2.2 相对指标的计算

随着研究目的和任务不同，对比的基数也不同，相对指标的就产生不同的种类和计算方法，以下介绍常用的几种。

1. 结构相对指标

结构相对指标是指在分组的基础上，将总体区分为性质不同的各个部分，以各部分指标数值与总体全部数值相比所得的相对数。它反映现象各部分在总体中所占的比例或比率，说明现象内部的组成状况与数量关系，其计算公式为

$$结构相对指标=\frac{总体中某一部分数值}{总体全部数值}\times 100\% \tag{4-1}$$

例如，2013 年年末，中国大陆总人口为 136 072 万人，其中男性人口为 69 728 万人，占人口总数的比例为 51.2%；女性人口为 66 344 万人，占人口总数的比例为 48.8%。这表明了我国人口的性别构成状况。

结构相对指标一般以百分数的形式表示。因为总体的全部数值等于总体内部各部分数值之和，所以总体内各部分占总体的比例(百分比)之和等于 100%或 1。应注意，结构相对指标的分子与分母位置不能互换。

结构相对指标在统计分析中应用非常广泛，常应用于消费结构分析。消费结构是指各类消费支出在总消费支出中所占的比例。19 世纪德国统计学家恩格尔(Engle)通过对英国、法国、德国、比利时等国居民家庭收支进行分析研究，提出了恩格尔定律，即随着家庭收入的增加，家庭收入或总支出中用于食品方面的支出比例越来越小。反映这个定律的结构相对数，成为恩格尔系数，其计算公式为

$$恩格尔系数=食品支出总额/消费支出总额 \tag{4-2}$$

将不同时期的结构相对指标进行对比，可以通过总体结构的变动，观察客观事物变化

的进程。一般地，事物的变化总是先从内部结构演变开始的，这种变化常常反映着事物发展由量变到质变的过程。掌握这一进程并加以分析，才能使我们认识事物发展的规律性。

【实例 4.1】 我国城镇和农村居民“十一五”期间家庭恩格尔系数指标见表 4－2。

表 4－2 我国“十一五”时期城乡居民家庭恩格尔系数

指标＼年份	2006	2007	2008	2009	2010
城镇居民家庭恩格尔系数	35.8%	36.3%	37.9%	36.5%	35.7%
农村居民家庭恩格尔系数	43.0%	43.1%	43.7%	41.0%	41.1%

（资料来源：国家统计局.）

从表 4－2 的数据可看出，“十一五”期间，我国农村居民家庭恩格尔系数下降趋势明显，城镇居民家庭恩格尔系数呈螺旋式下降态势。

2. 比例相对指标

比例相对指标是同一总体内不同组成部分之间的比值，反映各组成部分之间的联系和比例关系，其计算公式为

$$比例相对指标=\frac{总体中某部分数值}{总体中另一部分数值}\times 100\% \tag{4-3}$$

比例相对指标可以用百分数表示，也可以用一比几或几比几形式表示。比例相对指标的分子、分母同属一个总体，而且分子、分母的位置可以互换。

【实例 4.2】 2013 年年末中国大陆总人口为 136 072 万人，其中，城镇人口为 73 111 万人，乡村人口为 62 961 万人，则城镇人口与乡村人口的比例可表示为 1∶0.861 2。

【实例 4.3】 我国 2013 年国内生产总值 568 845 亿元(初步核算数据)。其中，第一产业增加值 56 957 亿元；第二产业增加值 249 684 亿元；第三产业增加值 262 204 亿元。第一、第二和第三产业增加值的比例为 10.0∶43.9∶46.1。

分析总体中若干部分的比例关系时，可采用连比形式(1∶m∶n)。为了能清楚地反映各部分之间的数量关系，连比组数不宜过多。

比例相对指标对于国民经济宏观调控具有重要意义。例如，人口性别比例、工农业总值中农轻重比例、积累与消费比例等。利用比例相对指标可以分析国民经济中各种比例关系，从而调整不合理的比例，促使社会主义市场经济稳步协调发展。

3. 比较相对指标

比较相对指标又称类比相对指标，是将两个同类指标做静态对比得出的相对指标，表明同类现象在不同空间(如在各国、各地、各单位)下的数量对比关系，其计算公式为

$$比较相对指标=\frac{某条件下某类指标数值}{另一条件下同类指标数值}\times 100\% \tag{4-4}$$

式中，分子与分母现象所属统计指标的含义、口径、计算方法和计量单位必须一致。比较相对指标一般用百分数或倍数表示。

【实例 4.4】 2011 年北京、上海两城市有关资料见表 4－3。



表 4-3 2011 年两城市相关资料比较

城 市	常住人口数/万人	国内生产总值/亿元	人均国内生产总值/（元/人）
北京	2 018.6	16 000.4	79 265
上海	2 347.46	19 195.69	81 772
比较相对指标（以北京为 100）	116.29	119.97	103.16

从表 4-3 的数据可看出，虽然上海市国内生产总值比北京市多 19.97%，但由于上海市的常住人口数比北京市多 16.29%，因此上海市的人均国内生产总值仅比北京市多 3.16%。

在实际工作中，计算比较相对指标时，作为比较基数的分母除了可以取常规的与分子分属不同空间的同类指标外，还可以将比较的基数典型化。例如，将本单位产品的质量、成本、单耗等各项技术经济指标都与国内外同行业的先进水平对比，与国家规定的质量标准对比等，从而找出差距，为提高本单位生产管理水平提供依据。

比较相对指标与比例相对指标类似，分子与分母也可以互换。两者的差别为比例相对指标是同一总体的不同部分比较，而比较相对指标是同类指标的不同空间比较；比例相对指标反映的比例关系有时是一种客观标准，违背这个标准，就会造成比例关系的失调，比较相对指标反映的是事物之间的对比关系，一般不存在比例正常或失调的关系。

4. 动态相对指标

动态相对指标又称为动态相对数或发展速度，是同类事物在不同时间上的指标之比，反映现象在时间上的发展变化方向和变化程度。其计算公式为

$$\text{动态相对指标}=\frac{\text{报告期指标数值}}{\text{基期指标数值}}\times 100\% \tag{4-5}$$

一般地，把用来作为对比基础的时期称为基期，而把同基期对比的时期称为报告期或计算期。

【实例 4.5】 2012 年我国农村居民人均纯收入 7 917 元，2011 年为 6 977 元，则

$$\text{动态相对指标}=\frac{7\ 917}{6\ 977}\times 100\%\approx 113.47\%$$

这表明 2012 年我国农村居民人均纯收入比 2011 年增长 13.47%。它反映了我国农村居民人均纯收入在两年间的变化情况。

动态相对指标将在本书第 5 章时间数列中的详细介绍。

5. 强度相对指标

强度相对指标又称为强度相对数，是两个性质不同但有一定联系的总量指标对比的结果，用来表明现象的强度、密度和普通程度的综合指标。其计算公式为

$$\text{强度相对指标}=\frac{\text{某一总量指标数值}}{\text{另一有联系而性质不同的总量指标数值}}\times 100\% \tag{4-6}$$

强度相对指标一般采用有名数形式。例如，我国的人均粮食产量、人均GDP、居民人均可支配收入，以及人均汽车拥有量等采用的是“千克/人”、“元/人”、“辆/万人”等复名数形式；商品流转速度用“次”或“天”表示，采用的是单名数形式。但也有采用无名数形式的。例如，人口自然增长率、商品流通费率、企业资金利润率等采用千分数、百分数等无名数形式。

【实例4.6】 2012年某市年平均人口数为900万人，面积为12 000平方千米，计算强度相对指标。

解：

$$该市人口密度=\frac{900}{1.2}=750(人/平方千米)$$

强度相对指标常带有“平均”的含义，但由于它的分子、分母分属两个不同总体，所以它与平均指标不同。同时，有些强度相对指标的分子与分母可以互换，因此，它有正指标和逆指标之分。强度相对指标数值的大小与现象的强度、密度和普遍程度成正比的是正指标；强度相对指标数值的大小与现象的强度、密度和普遍程度成反比的是逆指标。例如，在医疗卫生统计中，每千人拥有的医生数是正指标，每个医生所服务的人口数是逆指标。一般说来，正指标越大越好，逆指标则越小越好。

但要注意的是，在实际工作中，并不是所有的强度相对指标都有正指标和逆指标之分。例如，人口出生率、人口死亡率等指标的分子、分母是不能互换的。因此，在计算和应用强度相对指标时，应根据研究目的、使用习惯及说明问题的难易程度，选择使用正指标或逆指标。

6. 计划完成程度相对指标

计划完成程度相对指标是实际完成数与相应的计划数之比，用以表明计划完成的程度，也叫计划完成相对数或计划完成百分比(percentage of plan fulfilment)。计划完成程度相对指标是检查、监督计划执行情况的重要指标，也是正确评价工作成绩的重要依据。

计划完成程度相对指标一般用百分数表示，其基本计算公式为

$$计划完成程度相对指标=\frac{实际完成数}{计划数}\times100\% \tag{4-7}$$

式中，分母是原定的计划指标，分子则是计划执行过程中或执行过程结束后统计出来的实际完成数值。一般要求在指标含义、计算口径、计算方法、计量单位、空间范围等方面一致。但分子与分母也就是实际完成数与计划数所包含的时期长短可以是相同的，也可以不同，两者对比的意义有所区别。

(1) 计划完成程度的计算。在制订计划时，由于具体情况与计算要求不同，计划任务数的表现形式有绝对数、相对数、平均数3种。因此计算计划完成程度指标时，既可以用绝对数直接对比求得，也可以用相对数和平均数对比求得，具体计算方法如下：

① 计划指标为总量指标时，计算计划完成程度相对指标。这是计算计划完成程度指标的基本方法，具体计算公式为

$$计划完成程度相对指标=\frac{实际水平}{计划水平}\times100\% \tag{4-8}$$



【实例 4.7】 某销售部门2012年商品销售额计划为2 000万元，实际完成2 100万元，求该销售部门商品销售额计划完成情况。

解：

$$计划完成程度=\frac{2\ 100}{2\ 000}\times 100\%=105\%$$

计算结果表明，该销售部门商品销售额计划完成程度为105%，即超额完成5%(105%－1)，超额完成100万元(2 100万元－2 000万元)。

② 计划指标为相对指标时，计算计划完成程度相对指标。在实际经济工作中，有时计划指标是以相对指标形式规定的，即用提高或降低百分比规定计划任务，如劳动生产率、成本降低率、流通费用率等。在这种情况下，计算计划完成程度相对指标就需要以实际完成百分比与计划规定百分比进行对比，其计算公式为

$$\begin{aligned}计划完成程度相对指标&=\frac{实际完成百分比}{计划规定百分比}\times 100\%\\&=\frac{实际水平/基期水平}{计划水平/基期水平}\times 100\%\\&=\frac{实际水平}{计划水平}\times 100\%\end{aligned}$$

式中，计划规定百分比通常是以基期(如上年度)实际水平为基础确定的，也就是说，在规定计划任务时，将计划水平与基期实际水平对比求得计划提高或降低的百分比；实际完成百分比也以基期水平为基础，即以报告期实际水平与基期水平对比求得实际提高或降低的百分比。因此，上述公式可变形为

$$计划完成程度相对指标=\frac{1\pm 实际提高(降低)百分比}{1\pm 计划提高(降低)百分比}\times 100\% \tag{4-9}$$

可见，用总量指标和相对指标计算出来的指标结果是一致的。式(4-9)还表明：利用相对指标计算计划完成程度的指标时，不能用实际增长率(或降低率)除以计划增长率(或降低率)求得，而应包含原有基础(即以基期水平为100%)在内，这样计算才符合公式含义，计算结果才能与用总量指标计算出来的结果相一致。

【实例 4.8】 某鞋业公司计划规定劳动生产率比去年提高10%，实际提高了15%，求劳动生产率计划完成情况。

解：

$$\begin{aligned}劳动生产率计划完成程度&=\frac{1+150\%}{1+10\%}\times 100\%\\&=\frac{115\%}{110\%}\times 100\%\approx 104.5\%\end{aligned}$$

计算结果表明，该企业劳动生产率得到提高，超额4.5%完成计划。

【实例 4.9】 某鞋业公司产品成本计划规定降低4%，实际降低6%，求该企业成本降低率计划完成情况。

解：

$$成本降低率计划完成程度=\frac{1-6\%}{1-4\%}\times 100\%=\frac{94\%}{96\%}\times 100\%\approx 97.92\%$$

计算结果显示，计划完成百分比小于100%，因为它是成本降低计划，所以这同样表明该企业完成了计划，而且该企业产品实际成本比计划规定多降低2.08%(100%－97.92%＝2.08%)。

注意，在实际工作中，有时也用百分点描述计划完成情况。例如，在实例4.8中，劳动生产率实际比计划多提高了5个百分点(15%－10%＝5%)，但不能说成提高5%；在实例4.9中，公司的产品成本实际比计划降低了2个百分点(6%－4%＝2%)，但不能说成多降低2%。百分点是指以百分数形式表示的相对数指标的增减变动幅度或对比差额，是指被比较的相对数指标之间的增减量，而不是它们之间的比值。

③ 计划指标为平均指标时。计算计划完成程度相对指标。如果计划指标以平均指标的形式提出，如产品单位成本、员工平均工资等，这时可直接运用计划完成程度相对指标的基本公式，以实际平均水平与计划平均水平进行对比计算。

【实例4.10】 2012年红星自行车厂生产的自行车，产品单位成本计划为500元，实际为470元，求自行车产品单位成本计划完成情况。

解：

$$\text{单位成本计划完成程度}=\frac{470}{500}\times 100\%=94\%$$

计算结果表明，红星自行车厂自行车产品单位成本实际比计划降低了6%，单位成本降低了30元。

(2) 计算和分析长期计划完成情况的方法。由于长期计划(如5年计划)中所规定的指标性质不同，所以表示方法也不同，因此产生了两种不完全相同的考核长期计划执行情况的方法。

① 水平法。如果在长期计划中，只规定计划期最末一年应达到的水平，则在计划期中，哪一期达到了计划数，则该期就完成了计划。用水平法检查长期计划执行情况的计算公式为

$$\text{计划完成程度}=\frac{\text{长期计划末年实际达到的水平}}{\text{长期计划规定的末年水平}}\times 100\% \tag{4-10}$$

【实例4.11】 某公司“十一五”计划规定，公司在2010年应达到年产量200万吨水平，各年实际生产情况见表4－4，检查该公司产量“十一五”计划完成情况。

表4－4　某公司“十一五”期间实际产量

时　间	2006年	2007年	2008年	2009年				2010年			
				第一季度	第二季度	第三季度	第四季度	第一季度	第二季度	第三季度	第四季度
产量/万吨	143	159	168	35	40	45	55	60	62	68	70

解：

$$2010\text{年公司产量计划完成程度}=\frac{60+62+68+70}{200}\times 100\%=\frac{260}{200}\times 100\%=130\%$$

计算结果表明，该公司产量超额30%完成了“十一五”计划。那么，提前完成计划时间是多少呢？

公司在2006年、2007年、2008年、2009年的产量均未达到200万吨，而从2009年



的第二季度到2010年的第一季度的产量刚好达到200万吨(40+45+55+60=200)，说明到2010年的第一季度末完成了计划，提前了3个季度。

利用水平法考核长期计划执行情况，计算提前完成任务的时间时，是以“统计年”计算的。即只要有连续一年时间(可以跨年度)实际完成水平达到最后一年计划水平，就算完成了长期计划，余下的时间就是提前完成计划时间。

② 累计法。如果在长期计划中，计划指标规定的是整个计划期内累计应达到的水平，则采用累计法考核长期计划执行情况，即将计划期内各年实际完成累计数与长期计划规定的累计数进行对比检查计划完成情况。其计算公式为

$$计划完成程度=\frac{长期计划期间实际累计完成数}{长期计划规定的累计数}\times 100\% \tag{4-11}$$

【实例 4.12】 某地区“十一五”计划规定5年内基本建设投资总额为30亿元，实际执行结果见表4-5，检查该地区基本建设投资总额“十一五”计划完成情况。

表4-5 某地区“十一五”计划期间基本建设投资总额

时间	2006年	2007年	2008年	2009年		2010年			
				上半年	下半年	第一季度	第二季度	第三季度	第四季度
基本建设投资额/亿元	4.5	5.8	6.7	3.8	4.6	2.4	2.2	2.3	2.5

解：

$$该地区基本建设投资总额计划完成程度=\frac{34.8}{30}\times 100\%=116\%$$

结果表明，该地区基本建设投资额超额16%完成了“十一五”计划。那么，提前了多少时间完成计划呢?

该地区基本建设投资额从2006年起至2010年第二季度止，实际累计完成了30亿元(4.5+5.8+6.7+3.8+4.6+2.4+2.2=30)，说明该地区提前两个季度完成了基本建设投资额的“十一五”计划。

利用累计法考核长期计划执行情况时，计算提前完成计划的时间是很有意义的。将计划期的全部时间减去自计划执行之日起至累计实际完成数已达到计划指标的日期止，其剩余的时间为提前完成计划的时间。

(3) 计划执行进度的检查。在实际应用中，利用计划完成程度指标检查计划执行情况时，往往有两种情况。一种情况是检查计划执行的结果，带有总结的性质。这时公式中实际水平与计划水平所包含的时期长短是完全一致的。另一种情况是检查计划执行的进度和均衡程度，这时公式中实际水平与计划水平所包含的时期长短是不一样的。

【实例 4.13】 某公司所属3家企业2012年全年和第三季度计划产值与实际完成产值的资料见表4-6。

在表4-6中，(5)和(6)都表示计划完成程度指标，但各自说明不同的问题。(5)说明第三季度产值计划的完成情况：全公司第三季度超额完成计划3.3%；所属各企业完成程度不同，甲企业超额完成5.6%，丙企业则未完成计划。(6)说明到第三季度止，实际完成全年计划产值的进度情况：全公司完成75%，正好是全年计划的3/4，符合均衡生产的要

求；但各企业的完成情况并不均衡，丙企业未实现进度要求，只完成全年计划的68%，可见丙企业产值计划能否完成，是能否实现全公司全年计划的关键。

由以上分析可知，计划完成程度指标不仅可以检查计划执行的结果，更重要的是可以检查和分析计划执行的进度，反映计划执行的均衡性。

表4-6　某公司3家企业产值完成情况资料

企　业	计划总产值/10万元		实际完成总产值/10万元		第三季度总产值计划完成程度/%	到第三季度止年总产值计划完成程度/%
	全年	其中第三季度	第三季度	累计到第三季度止		
	(1)	(2)	(3)	(4)	(5)=(3)/(2)	(6)=(4)/(1)
甲	5 000	1 250	1 320	3 800	105.6	76.0
乙	1 800	480	480	1 349	100.0	74.9
丙	700	175	168	476	96.0	68.0
合计或平均	7 500	1 905	1 968	5 625	103.3	75.0

4.2.3　相对指标的应用

以上所叙述的各种相对指标，从不同角度反映了现象之间的联系。要正确计算和运用这些相对指标，发挥它们各自的作用，在统计实际中还应注意以下问题。

1. 对比指标的可比性原则

相对指标是通过相互联系的统计指标之间的对比，研究和分析客观现象之间的数量对比关系的，因此保持对比指标之间的可比性，是运用相对指标的基本原则。客观内容不可比、统计范围或者统计口径不可比、计算方法和计量单位不可比，均会导致统计分析结论的谬误。在实际工作中，可比性是计算和运用相对指标的重要原则。可比性主要包括以下两个方面。

(1) 总体范围可比。例如，在新的一年里，某公司兼并了同行业另一家公司，企业规模扩大了，若计算产量动态相对指标，由于总体范围的变化，两年产量不可直接对比。另外，相对指标的分子与分母在范围上要相互适应。例如，文盲和半文盲占总人口的比例，按规定文盲和半文盲是指12周岁以上不识字或识字很少的人，这样在分母的总人口中，就应将12周岁以下的人口扣除才比较合理。

(2) 计量单位和计算方法可比。这是将同类现象在不同空间和时间上进行对比时应注意的问题。例如，对比两个不同时期的社会总产值，就要考虑是按现行价格计算还是按不变价格计算。因为只有扣除价格变动因素，才能真正比较出不同时期产值的变化。如果是在国际间进行两个指标的对比，就更应慎重考虑指标的范围、内容和计算方法是否具有可比性。例如，西方国家的国民收入与我国国民收入指标的经济含义、所包括的范围及计算方法都有所不同，不能直接进行对比。



遇到分子与分母不可比时，可以进行调整以使它们可比。例如，不同时期的行政区划、组织机构或隶属关系有所不同，这时应以报告期为准，对基期数值进行调整。又如，指标计算方法、范围和口径的改变，使指标数值不能直接对比，这时也需将基期数据按报告期口径进行调整，以使两指标变为可比。分子与分母的计量单位不同时，也应先将其化为统一，才可比较。

2. 相对指标与统计分组法的结合运用

统计分组法是统计分析中经常运用的重要方法，结构相对指标、比例相对指标等必须以其为基础加以运用。除此之外，为了全面地分析客观现象的计划完成情况、发展变化关系，同样必须运用科学的分组方法。因此，相对指标与统计分组的科学配合是统计研究的一项重要内容。

3. 各种相互有联系的相对指标的结合运用

在实际工作中，把相互有联系的各种相对指标结合起来运用，可以更深刻、更全面地说明客观事物的性质，以得出正确结论。例如，在“做一做”中，为描述2012年公司取得的辉煌业绩，可以用2012年公司实际产量与2011年、2010年产量相比，反映动态发展趋势；再用2012年公司产量与同行业的竞争对手PROPER公司相比较，反映差距或实力；同时用2012年公司实际产量与公司的计划产量相比较，反映计划完成情况；还要用2012年产量与该公司“十二五”计划总产量相比所得的比例，反映计划完成进度。几种相对指标联系起来观察，才可以全方位、多角度、生动、丰富地描述公司业绩，进行科学评价。

4. 总量指标与相对指标的结合运用

由于总量指标和相对指标的作用不同，为了能获得对社会经济现象比较完整和深入的认识，需要将二者结合起来运用。

相对指标与总量指标相比，能更明显反映现象之间的联系和对比关系。但相对指标一般是由两个总量指标对比形成的，其比值是把总量指标(即绝对数)抽象化之后的结果。因此，仅从相对指标上看不出现象原有的规模和水平，这就掩盖了现象之间绝对量上的差别。由于在事物的比较过程中，有时候会出现总量指标较大、相对指标较小，或总量指标较小、相对指标较大的现象。因此，在运用相对指标研究客观现象的数量对比关系时，必须联系其所依据的基本指标，即必须与其背后的绝对水平和绝对差异相结合，以便客观地评价事物的发展变化情况。例如，2012年广州城市居民人均可支配收入38 054元，贵阳为21 796元，若两城市的居民人均可支配收入在2013年均增加10%，从相对水平来看增幅相同，但结合绝对水平来看，广州是在基期的38 054元基础上增长10%，增长的绝对量为3 805.4元，而贵阳是在基期21 796元的基础上增长10%，增长的绝对量是2 179.6元。从增加的实际效果看，广州是贵阳的1.75倍多。所以，只有把相对指标与总量指标结合运用，才能更全面地分析和说明问题。

4.3 平均指标

4.3.1 平均指标的意义和种类

平均指标又称平均数，反映现象总体各单位某一数量标志值的典型水平、一般水平和代表性水平。平均指标是社会经济现象中最常用的一种综合指标。

例如，我们经常计算的某门课程的班级平均分数，就是全班同学成绩的一般水平。在全班同学总体中，每位同学是总体单位，成绩80分、90分、78分等都是数量标志值。大家成绩参差不齐，各不相同，如何反映全班同学总体水平？经过计算，用平均分数(假定80分)反映。平均分数是全班同学成绩的最一般水平、最典型水平，最具有代表性。

平均指标的显著特点：它不是某一单位的具体数值，而是代表总体某种数量标志的一般水平，是总体各单位的代表值；平均指标把总体各单位标志值的差异给抽象化了，它是一个抽象化的数值。正是由于平均指标的“抽象化”特征，当计算出某地平均每户人口规模2.94人时，不必对数值进行四舍五入。

由于平均指标能够综合反映某种社会经济现象总体在一定条件下的一般水平，所以它的应用很广泛，其作用主要表现在以下几个方面。

1. 可以反映总体各单位分布的集中趋势

在社会经济现象总体中，各单位中某一数量标志值往往各不相同。通常情况下，标志值很大或标志值很小的单位比较少，而逐渐靠近平均数的单位数逐渐增多，标志值围绕在平均数周围的单位占最大比例。平均指标反映了标志值变动的集中趋势。例如，某市职工年平均工资45 800元就反映了职工工资水平的集中趋势，是职工工资在具体条件下所达到的一般水平。

2. 可以对现象在不同空间、时间上进行比较分析

对于不同国家、不同地区、不同单位的同类现象的水平，由于总体范围的大小可能不同，通常不能直接进行比较分析，然而通过计算平均指标，可以使现象具有可比性。例如，甲乙两家公司12月职工工资总额分别是300万元和600万元，职工人数分别为800人和2 000人，尽管从总量指标“工资总额”和“职工人数”看，乙公司均高于甲公司，但从“人均工资”这一平均指标看，甲公司要高于乙公司750元/人。这个结果表明，甲公司工资水平高于乙公司。这个指标是通过平均指标的空间对比获得的。

根据同样道理，利用平均指标可以对同一现象在不同时间进行比较分析。例如，将现在职工平均工资与改革开放前的工资水平进行比较分析，从中显示出职工工资水平的大幅度提高。

3. 可以分析现象之间的依存关系

由于平均指标是代表总体某种数量标志的一般水平，在对现象总体进行分组的基础



上，可以通过计算各组的平均指标，分析现象之间的依存关系。

4. 可以作为评价事物的参考依据

事物的好坏是相对的，通常以代表事物发展一般水平的平均指标为评定依据。例如，考生成绩为78分，如何判断成绩的高低？若平均成绩为60分，则78分比较高；若平均成绩为90分，则78分比较低。

5. 可以进行数量上的估算

在统计估计和预测中，往往利用并根据部分单位标志值计算的平均指标推算总体平均指标，或者以总平均指标推算总体标志总量。例如，已知市内某区牛奶的人均消费量，可以采用科学方法，推断全市的人均牛奶消费量，也可以推断全市牛奶的消费总量。

平均指标按其所属总体的时间范围不同分为两种：静态平均数和动态平均数。静态平均数是反映同一时间范围内总体各单位某一数量标志一般水平的平均指标；动态平均数是反映不同时间而同一空间范围内总体某一数量标志一般水平的平均指标。本章主要讨论静态平均数，动态平均数将在第5章专门讨论。

静态平均数按其计算方法的不同分为两种：数值平均数和位置平均数。

根据总体各单位标志值计算的平均数，称为数值平均数，常见的主要有算术平均数(arithmetic mean)、调和平均数(harmonic mean)和几何平均数(geometric mean)等；根据总体标志值在分配数列中的位置确定的平均数，称为位置平均数，常见的主要有众数(mode)、中位数(median)和四分位数(quartiles)等。

4.3.2 数值平均数

1. 算术平均数

1）算术平均数的意义

算术平均数是总体标志总量与总体单位总量对比的结果，其基本公式为

$$\text{算术平均数}=\frac{\text{总体标志总量}}{\text{总体单位总量}} \tag{4-12}$$

算术平均数与强度相对指标都是比值，都有“平均”的含义，但两者明显区别在于：算术平均数的分子和分母是同一个总体的两个总量指标，分子是标志总量，分母是单位总量，而且分子、分母位置不能互换；强度相对指标的分子和分母分属两个不同总体的总量指标，而且分子、分母位置颠倒有意义，它有正、逆指标之分。例如，一个三口之家，父亲和母亲均在岗，每月两人工资总额为9 000元，孩子上小学。则9 000元除以2，表示夫妻平均工资4 500元/人，9 000元除以3，表示全家人均收入3 000元/人。在这里，孩子没有收入，但分母单位总量里包含他，则分母总体扩大，与分子是不同的两个总体，所以，“人均收入”是强度相对指标，而“平均工资”才是同一总体的标志总量与单位总量之比，是平均指标。

算术平均数是分析社会经济现象一般水平和典型特征的最基本、最常用的一种平均指

标。它也是平均指标中最重要的一种。一般不加特别说明时，所称的“平均数”都是指算术平均数。

2）算术平均数的分类

根据基本公式计算算术平均数时，由于依据的资料不同，计算方法有所不同，可分为简单算术平均数和加权算术平均数。

（1）简单算术平均数。

如果总体单位数不多，资料未经分组，计算算术平均数时可采用简单算术平均数的方法。计算过程是先把总体各单位标志值相加得出总体标志总量，再与总体单位总数对比，即可得到算术平均数。计算公式为

$$\bar{x}=\frac{x_1+x_2+x_3+\cdots+x_n}{n}=\frac{\sum x}{n} \tag{4-13}$$

式中：$\bar{x}$ 为算术平均数；x 为总体各单位标志值；n 为总体单位数；$\sum$ 为加总的符号。

【实例 4.14】 一个公司有 5 个部门，每个部门员工数分别为 24、13、19、26 和 11，求平均每部门的人数。

解：

$$\text{平均人数}=\frac{24+13+19+26+11}{5}=\frac{93}{5}=18.6(\text{人})$$

（2）加权算术平均数。

如果调查所得的原始资料已经经过分组整理，形成了变量数列，则计算算术平均数要采用加权算术平均数的方法。计算过程是将各组的变量值与各组的单位数相乘，计算出各组标志总量，将各组标志总量汇总得出总体标志总量，然后除以各组单位数之和即总体单位总量，得到平均数。计算公式为

$$\bar{x}=\frac{x_1f_1+x_2f_2+x_3f_3+\cdots+x_nf_n}{f_1+f_2+f_3+\cdots+f_n}=\frac{\sum xf}{\sum f} \tag{4-14}$$

式中：x 为各组标志值；f 为各组次数。

【实例 4.15】 服装商店要销售 100 件毛衣，其中 20 件大号毛衣，每件 200 元；50 件中号毛衣，每件 190 元；30 件小号毛衣，每件 180 元。计算毛衣的平均价格。

解：根据题意，可列出计算结果见表 4-7。

表 4-7　服装商店毛衣销售价格资料及其计算

销售价格(x)/元	件数(f)/件	销售总价值(xf)/元	销售价格(x)/元	件数(f)/件	销售总价值(xf)/元
200	20	4 000	180	30	5 400
190	50	9 500	合计	100	18 900

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{18\ 900}{100}=189(\text{元})$$

由上式可见，平均数的大小不仅取决于总体各单位标志值 x，同时也取决于各标志值的次数 f。次数多的标志值对平均数的影响要大些，次数少的标志值对平均数的影响相应



地小些。标志值次数的多少，对平均值的大小有权衡轻重的影响作用，所以称其为权数。

权数不但可以用次数、频数(即总体各组单位数)这种绝对数表示，还可以用比例、频率这种相对数表示。此时，加权算术平均数公式可以演化为

$$\bar{x}=\sum x\frac{f}{\sum f} \tag{4-15}$$

仍以表4-7中的资料为例，各型号毛衣价格及件数构成资料见表4-8。

表4-8 服装商店毛衣销售价格资料及其计算

销售价格(x)/元	件数构成$\left(\frac{f}{\sum f}\right)$	$x\frac{f}{\sum f}$	销售价格(x)/元	件数构成$\left(\frac{f}{\sum f}\right)$	$x\frac{f}{\sum f}$
200	20%	40	180	30%	54
190	50%	95	合计	100%	189

解：

$$\bar{x}=\sum x\cdot\frac{f}{\sum f}=189(\text{元})$$

由上例明显看出，比例的大小直接表明了该组标志值占据平均数的地位。所以，权数对于算术平均数的影响作用，不是决定于权数本身数值的大小，而是决定于权数比例(或称为相对权数)的大小。权数比例是指作为权数的各组单位数占总体单位数的比例。哪一组的单位数所占的比例大，哪一组标志值对平均数的影响就大。

当各组的单位数相等或各组单位数所占的比例相等时，权数对各组的作用都一样，加权也就失去了意义。此时，加权算术平均数等于简单算术平均数，即当$f_1=f_2=f_3=\cdots=f_n$时，$\bar{x}=\frac{\sum xf}{\sum f}=\frac{\sum x}{n}$。可见，简单算术平均数是加权算术平均数的特例。

在计算加权算术平均数时，还会遇到权数的选择问题。选择权数的原则是，务必使各组的标志值与其乘积等于各组的标志总量，并且具有实际的经济意义。在分配数列条件下，一般来说，次数就是权数。但也有例外，特别是用相对数或平均数计算加权算术平均数时，要特别注意。

【实例4.16】 某居住小区60户居民人均居住面积资料见表4-9，求总人均居住面积。

表4-9 小区居民人均居住面积资料及其计算

人均居住面积(x)/(平方米/人)	户数/户	居住人数(f)/人	居住面积(xf)/平方米
60	3	6	360
40	47	141	5 640
30	4	16	480
24	6	30	720
合计	60	193	7 200

本例要平均的对象是各户的人均居住面积，为计算小区总人均居住面积，应该以居住

人数为权数，户数虽是不同人均居住面积的次数但并不是合适的权数。因为人均居住面积与居住人数相乘才有具体的经济含义，即为居住总面积。小区总人均居住面积计算结果如下：

解：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{7\ 200}{193}\approx 37.31(\text{平方米/人})$$

上面的例题采用的都是单项式数列，如果是组距式分配数列，则先计算组中值来代表各组的标志值 x，再计算平均数。但应用这种计算方法需有一个假定条件，即假定各单位标志值在各组内是均匀分布的，而实际上，这是很难达到的。所以，这种用组中值代替计算的算术平均数，不可避免地会存在一定程度的误差，具有近似值的性质。

【实例 4.17】 某班学生“统计学基础”测验成绩被编制成组距式分配数列见表 4－10 所示，计算班级平均成绩。

表 4－10 “统计学基础”测验成绩资料及其计算

成绩/分	学生数(*f*)	组中值(*x*)	*xf*	成绩/分	学生数(*f*)	组中值(*x*)	*xf*
60 以下	2	55	110	80～90	10	85	850
60～70	12	65	780	90 以上	5	95	475
70～80	21	75	1 575	合计	50	—	3 790

解：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{3\ 790}{50}=75.8(\text{分})$$

同样的道理，在“做一做”中，根据 BLOOMY 公司 2012 年下属企业生产情况，则可计算结果见表 4－11。

表 4－11 BLOOMY 公司 2012 年下属企业生产情况计算

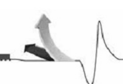
按产量计划完成程度分组	企业数/家	计划产量(*f*)/万台	组中值(*x*)	*xf*
100%～110%	2	13	105%	13.65
110%～120%	7	53	115%	60.95
120%～130%	3	24	125%	30
—	12	90	—	104.6

则 2012 年全公司平均计划完成程度 $\bar{x}=\frac{\sum xf}{\sum f}\times 100\%=104.6/90\approx 116.22\%$

2. 调和平均数

1） 简单调和平均数

简单调和平均数是各个标志值倒数的简单算术平均数的倒数。其计算公式为



$$\bar{x}_h = \frac{n}{\frac{1}{x_1}+\frac{1}{x_n}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}} \tag{4-16}$$

式中：$\overline{x_h}$ 为调和平均数；x 为各标志值；n 为项数。

【实例 4.18】 3 个蔬菜超市销售同一种蔬菜，但价格不同，每千克蔬菜价格分别为 1.8 元、2 元、2.3 元。若在 3 个超市各购买 1 元的这种蔬菜，则蔬菜的平均价格为多少？

解：

$$平均价格 = \frac{n}{\sum \frac{1}{x}} = \frac{1+1+1}{\frac{1}{1.8}+\frac{1}{2}+\frac{1}{2.3}} \approx 2.01(元)$$

上式中，蔬菜平均价格的计算采用了新的计算方法，因为给出的资料与计算算术平均数的资料不同。在这里，有各组的标志值即价格，没有各组的次数(购买量)，却有两者的乘积(购买额)，所以，分母购买量利用购买额除以价格得到。可见，调和平均数与算术平均数计算原理相同，有观点说，调和平均数是算术平均数的变形。

2）加权调和平均数

加权调和平均数是各个标志值倒数的加权算术平均数的倒数。在实例 4.18 中，各组购买额恰好相等，均为 1 元，如果购买额不相等，则必须采用加权调和平均数。其计算公式为

$$\overline{x_h} = \frac{m_1+m_2+m_3+\cdots+m_n}{\frac{m_1}{x_1}+\frac{m_2}{x_2}+\frac{m_3}{x_3}+\cdots+\frac{m_n}{x_n}} = \frac{\sum m}{\sum \frac{m}{x}} \tag{4-17}$$

式中：m 为总体各组标志总量。

【实例 4.19】 学校食堂购进某种蔬菜，相关资料见表 4－12，求蔬菜的平均价格。

表 4－12 蔬菜价格资料及其计算

购买地点	价格(x)/(元/千克)	购买金额(m)/元	购买量(m/x)/千克
甲超市	1.8	20	11.11
乙超市	2	15	7.5
丙超市	2.3	10	4.35
合计	—	45	22.96

解：

$$蔬菜的平均价格\ \overline{x_h} = \frac{\sum m}{\sum \frac{m}{x}} = \frac{45}{22.96} \approx 1.96(元)$$

调和平均数与算术平均数在计算上是相通的，但各自适合不同的资料。在实际应用中，一般采用下列 5 个步骤进行计算。

（1）在给出的材料中找出 x。方法是“求谁的平均数，谁就是 x”。

（2）确定权数为 f 还是为 m。方法是分配数列中的这栏数值若与 x 相乘有意义，则为 f；若与 x 相除有意义，则为 m。

(3) 写出计算公式。若权数定为 f，公式为 $\bar{x}=\frac{\sum xf}{\sum f}$，即采用算术平均数；若权数定为 m，则公式为 $\bar{x}=\frac{\sum m}{\sum \frac{m}{x}}$，即采用调和平均数。

(4) 延展计算表。根据写出的公式，需要什么内容就在表格中列出相应内容。例如，若采用的是算术平均数公式，则表格中列出 xf 以便于计算；若采用的是调和平均数公式，则表格中列出 $\frac{m}{x}$ 以便于计算。

(5) 将数值代入公式，进行计算，得到答案。

3. 几何平均数

几何平均数是若干项变量值的连乘积开若干次项数的方根。它是计算平均数的另一种形式，主要用于计算比率或速度的平均。当所掌握的变量值本身是比率的形式，而且各比率的乘积等于总的比率时，就采用几何平均法计算平均比率。因为这与几何平均数的数学性质相一致。根据所掌握的资料不同，几何平均数分为简单几何平均数和加权几何平均数。

1) 简单几何平均数

如果资料未分组，直接将 n 项变量值连乘，然后对其连乘积开 n 次方所得的平均数为简单几何平均数。其计算公式为

$$\overline{x_G}=\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\cdots\cdot x_n}=\sqrt[n]{\prod x} \tag{4-18}$$

式中：$\overline{x_G}$ 为几何平均数；x 为各标志值；n 为变量值个数；$\prod$ 为连乘符号。

【实例 4.20】 产品的生产往往需要几道生产工序，只有在第一道工序合格的产品才能进入第二道工序。现已知纺织厂纺纱车间产品合格率为 91%，织布车间产品合格率为 89%，印染车间产品合格率为 87%，求 3 个车间平均产品合格率。

解：

由于后续车间的合格率是在前一车间产品合格基础上计算的，所以全厂产品的总合格率并不等于各车间产品合格率的总和，而是各车间产品合格率的连乘积，因此要采用几何平均法计算各车间产品平均合格率，即产品平均合格率为

$$x_G=\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\cdots\cdot x_n}=\sqrt[3]{91\%\times 89\%\times 87\%}\approx 88.99\%$$

2) 加权几何平均数

与算术平均数一样，当资料已经经过整理，应以各变量值出现的次数为权数，计算加权几何平均数。其计算公式为

$$\overline{x_G}=\sqrt[\sum f]{x_1^{f_1}\cdot x_2^{f_2}\cdot x_3^{f_3}\cdot\cdots\cdot x_n^{f_n}} \tag{4-19}$$

式中：f 为各变量值次数；$\sum f$ 为次数总和。

【实例 4.21】 一条产品流水线由 12 道工序组成，其中，合格率为 98%的有 2 道工序，合格率为 96%的有 5 道工序，合格率为 92%的有 3 道工序，合格率为 89%的有 2 道工序。



求产品总平均合格率。

解：

产品总平均合格率为

$$\overline{x_G}=\sqrt[12]{98\%^2\times96\%^5\times92\%^3\times89\%^2}\times100\%\approx94.12\%$$

4.3.3 位置平均数

数值平均数都是根据标志值计算得到的，而位置平均数是根据总体各单位标志值所处的位置确定的。位置平均数主要包括众数、中位数和四分位数。

1. 众数

1）众数的意义

众数是指总体中出现次数最多的标志值，是总体各单位一般水平的代表值，反映现象的集中趋势，用 M_o表示。众数可能不存在或不唯一。

在实际工作中，众数的应用比较广泛。例如，为掌握消费者需要的鞋的号码大小，不需要全面登记所有鞋码进行平均，只用生活中最普遍的、成交量最大的尺码，它反映了人们一般的需求。

2）众数的确定

众数的确定方法根据具体资料而定。

（1）由未分组资料确定众数。在资料未分组情况下，众数的确定很简单，只需找出次数最多的标志值即可。例如，一组学生年龄(单位：岁)分别为 18、19、19、20、20、20、20、22，则众数为 20；若学生年龄为 18、19、19、19、20、20、20、22，则有双众数，分别是 19、20；若学生年龄为 16、17、18、19、20、21、22、23，则不存在众数。

（2）由单项式数列确定众数。在单项式数列情况下，次数最多的组的标志值便是众数，见表 4－13。

表 4－13 某商场销售成年女鞋资料

女鞋号码/码	销售量/百双	女鞋号码/码	销售量/百双
35	1	39	0.6
36	2	40	0.2
37	5	合计	10.9
38	2.1		

从表 4－13 中可以看出，出现次数最多(5 百双)的变量值是 37 码的鞋，因此，37 码就是该商场女鞋销售的众数。

（3）由组距数列确定众数。在组距数列情况下，计算众数方法：首先在数列中找到次数最多的组；然后用公式计算众数的近似值。相关计算公式为

$$\text{下限公式：}M_o=L+\frac{\Delta_1}{\Delta_1+\Delta_2}\times d \tag{4-20}$$

$$\text{上限公式：}M_o = U - \frac{\Delta_2}{\Delta_1 + \Delta_2} \times d \quad (4\text{-}21)$$

式中：M_o为众数；L 为众数所在组的下限；U 为众数所在组的上限；Δ_1 为众数所在组次数与前一组次数之差；Δ_2 为众数所在组次数与后一组次数之差；d 为众数所在组的组距。

【实例 4.22】 某班 30 名大学生的体重资料见表 4－14，求 30 名大学生的体重的众数。

表 4－14　某班 30 名大学生的体重资料

体重/千克	人数/人	体重/千克	人数/人
44～48	2	60～64	7
48～52	3	64～68	2
52～56	6	合计	30
56～60	10		

解：

为求众数，首先确定数列的众数组。实例 4.22 中次数最多的组是“56～60”组，按照下限公式，众数值近似确定为

$$M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times d = 56 + \frac{4}{4+3} \times 4 \approx 58.29(\text{千克})$$

若将资料代入上限公式，计算结果与下限公式结果一样。

3）众数的特点及应用

（1）众数是根据变量值出现的次数确定的，而不是通过所有变量值计算得到的，所以，众数不受极端变量值的影响。众数的这一特点是数值平均数所不具备的。在实际工作中，众数用得最多的是具有明显偏态集中趋势的次数分配。例如，按照统计国际惯例，对家庭收入分配数列、工人周工资分配数列、某种债券息票率分组的行情次数等进行的分析，都采用出现次数最多的众数，得到“最普通的家庭收入额”、“最普通的工人周工资金额”、“最常见的外汇率、息票率”等。

（2）众数是出现次数最多的变量值，如果数据分布没有明显集中趋势，众数就可能不存在；如果有两个最高次数，也可以有两个众数(bimodal)。只有在总体单位比较多，而且又明显地集中于某个变量值时，众数计算才有意义。

（3）众数主要用于测度定类数据的集中趋势，当然也适用于作为定序数据，以及定距和定比数据集中趋势的测度值。众数是唯一一个能用在名词数据上的平均数。例如，一项对大学生的研究包括了 10 个心理学专业的学生，20 个英语专业的学生和 6 个管理专业的学生。我们无法计算这些专业的平均数，但可以指出众数是英语专业，因为它是出现次数最多的专业。在这里，“英语专业”是定类数据的集中趋势。

2. 中位数和四分位数

1）中位数的意义

中位数是指将总体各单位标志值按一定顺序排列后，处于中间位置的标志值。由于它的位置居中，其数值不受极端数值的影响，也能表明总体标志值的一般水平，用 M_e 表示。



中位数将全部标志值分成两半，一半小于中位数，一半大于中位数，所以中位数又称为二分位数。

在实际工作中，有许多场合用中位数表示现象的一般水平。例如，在研究居民收入水平时，以居民收入中位数代表居民收入水平比采用算术平均数进行计算更为科学。

2）中位数的确定

中位数的确定方法根据具体资料而定。

（1）由未分组资料确定中位数在资料未分组的情况下，中位数的确定比较简单。首先，把标志值按大小顺序排列起来，然后，计算中位数所在的位置找出中位数。如果总体单位数是奇数，处于 $\frac{n+1}{2}$（n 代表总体单位数）位置的标志值就是中位数；如果总体单位数是偶数，那么中位数就是处于 $\frac{n}{2}$ 与 $\frac{n}{2}+1$ 位置的两个标志值的算术平均数。

例如，某班甲、乙两个小组各有 7 人和 8 人，学生身高(单位：厘米)资料如下：

甲组学生身高：154，160，162，164，165，166，168；

乙组学生身高：157，159，162，163，164，167，168，172。

甲组中，中位数位置是$\frac{7+1}{2}=4$，中位数是 164 厘米。

乙组中，中位数位置是$\frac{8}{2}=4$与$\frac{8}{2}+1=5$之间，中位数是$\frac{163+164}{2}=163.5$(厘米)。

（2）由单项式数列确定中位数。由单项式数列确定中位数时，首先将各组次数累计，然后根据公式 $\frac{\sum f+1}{2}$ 确定中位数的位次，最后从次数累计中找到中位数所在的组，该组的标志值就是中位数。

【实例 4.23】 某学院电子商务专业两个班学生参加某项技能测试，在规定的时间内完成任务数各不相同，资料见表 4-15，求每班学生完成任务个数的中位数。

表 4-15 电子商务专业两个班学生某项技能测试完成任务个数资料

按完成任务数分组/个	一班学生数/人	二班学生数/人	一班人数累计/人	二班人数累计/人
1	3	4	3	4
2	5	5	8	9
3	8	6	16	15
4	7	8	23	23
5	5	7	28	30
合计	28	30	—	—

解：

一班学生完成任务数的中位数位次是 $\frac{\sum f+1}{2}=\frac{28+1}{2}=14.5$，说明中位数位于第 14 和第 15 个学生之间。由累计次数知，第 14 和第 15 个学生都在第三组，标志值均为 3，



所以，一班学生完成任务个数的中位数是 3。

二班学生完成任务数的中位数位次是 $\frac{\sum f+1}{2}=\frac{30+1}{2}=15.5$，说明中位数位于第 15 和第 16 个学生之间。由累计次数知，第 15 和第 16 个学生分属第三组和第四组，所以应取这两组标志值的算术平均数作为中位数。所以，二班学生完成任务个数的中位数是 $\frac{3+4}{2}=3.5$。

(3) 由组距数列确定中位数。在组距数列情况下，计算中位数的方法：首先，根据累计次数 $\frac{\sum f+1}{2}$ 确定中位数所在组，这个组的上下限就规定了中位数可能的取值范围；然后，用公式计算中位数的近似值。相关计算公式为

$$\text{下限公式：} M_e = L+\frac{\frac{\sum f}{2}-S_{m-1}}{f_m}\times d \tag{4-22}$$

$$\text{上限公式：} M_e = U-\frac{\frac{\sum f}{2}-S_{m-1}}{f_m}\times d \tag{4-23}$$

式中：M_e 为中位数；L 为中位数所在组的下限；U 为中位数所在组的上限；f_m 为中位数所在组的次数；S_{m-1} 为中位数所在组以前各组的累计次数；S_{m+1} 为中位数所在组以后各组的累计次数；$\sum f$ 为总次数；d 为中位数所在组的组距。

【实例 4.24】 根据表 4-14 给出的某班 30 名大学生的体重资料，求大学生体重的中位数。

解：

为求中位数，首先应求出中位数所在组，在总人数 30 的前提下，中位数所在组为“56～60”组，按照下限公式，中位数值近似确定为

$$M_e = U+\frac{\frac{\sum f}{2}-S_{m-1}}{f_m}\times d = 56+\frac{\frac{30}{2}-11}{10}\times 4 = 57.6(\text{千克})$$

若将资料代入上限公式，计算结果与下限公式结果一样。

3) 中位数的特点及应用

(1) 中位数是一种位置平均数，其大小取决于它在序列中的位置，因此它不受极端数值的影响。当存在极端数值时，中位数能比数值平均数更好地代表数据分布的一般水平。

(2) 中位数处于中间位置，有一半数值小于中位数，另有一半数值大于中位数，所以，它能表明数字资料的集中趋势。在实际工作中，中位数用得较多的是农村、城镇居民人均可支配收入及测定人口年龄分配的平均年龄数。例如，2012 年，我国农村居民人均纯收入中位数为 7 019 元；城镇居民人均可支配收入中位数为 21 986 元。按照统计国际惯例，各国政府统计工作对年龄分组采用中位数，而不用算术平均数。因为人口的年龄分配不是中间高、两边低的分配形态，而是 J 形的分配形态。婴儿 0～1 岁的人数最多，随着年龄增大，人数逐渐减少，到百岁左右所剩人数很少。若计算算术平均数，则会由于老人



岁数很大的影响，使得算术平均年龄偏大，从而与实际情况不符。

(3) 中位数主要用于测度定序数据的集中趋势，当然也适用于定距数距和定比数据的集中趋势，但不适用于定类数据。

4) 四分位数

中位数是从中间点将全部数据等分为两部分，也称二分位数。同样，可通过3个数值，将全部标志值分割为4个相等部分，其中每部分包含25%的数据，处在分位点上的这3个分割的数值就是四分位数，分别记为 Q_1、Q_2 和 Q_3。很显然，中间的四分位数就是中位数，因此通常所说的四分位数是指第一个四分位数 Q_1 和第三个四分位数 Q_3。

四分位数的确定方法与中位数类似，也要分为两种情况，即未分组资料和分组资料。对于未分组资料，首先将数据按大小顺序排序，然后计算分位数所在的位置，最后确定相关的四分位数。

【实例 4.25】 某公司11位员工6月生产某种产品的数量(单位：千克)分别是17、16.8、16.6、16.9、18.4、17.7、18、18.2、17.5、16、18.6，计算四分位数。

解：

将数据按大小顺序排列成

16　16.6　16.8　16.9　17　17.5　17.7　18　18.2　18.4　18.6

分位数所在的位置分别为

$$Q_1\text{ 的位置}=\frac{n+1}{4}=\frac{11+1}{4}=3$$

$$Q_2\text{ 的位置}=\frac{2(n+1)}{4}=\frac{n+1}{2}=\frac{11+1}{2}=6$$

$$Q_3\text{ 的位置}=\frac{3(n+1)}{4}=\frac{3(11+1)}{4}=9$$

因此，四分位数分别为 $Q_1=16.8$(千克)，$Q_2=17.5$(千克)，$Q_3=18.2$(千克)。

实例4.25中的资料项数 n 加上1后恰好为4的倍数，所以，四分位数的确定很简单。若 $n+1$ 不是4的整数倍数，按上述公式确定的四分位数的位置就带有小数，这时有关的四分位数就应该是与该小数相邻的两个整数位置上的标志值的加权算术平均数，权数的大小取决于两个整数位置距离的远近。距离越近，权数越大；距离越远，权数越小，权数之和等于1。

【实例 4.26】 某公司10位员工6月生产某种产品的数量(单位：千克)分别是17、16.8、16.6、16.9、18.4、17.7、18、18.2、17.5、16，计算四分位数。

解：

将数据按大小顺序排列成

16　16.6　16.8　16.9　17　17.5　17.7　18　18.2　18.4

四分位数所在的位置分别为

$$Q_1\text{ 的位置}=\frac{n+1}{4}=\frac{10+1}{4}=2.75$$

$$Q_2\text{ 的位置}=\frac{2(n+1)}{4}=\frac{n+1}{2}=\frac{10+1}{2}=5.5$$

$$Q_3\text{ 的位置}=\frac{3(n+1)}{4}=\frac{3(10+1)}{4}=8.25$$

则四分位数分别为

$$Q_1=0.25\times\text{第 2 项}+0.75\times\text{第 3 项}=0.25\times16.6+0.75\times16.8=16.75(\text{千克})$$

$$Q_2=0.5\times\text{第 5 项}+0.5\times\text{第 6 项}=0.5\times17+0.5\times17.5=17.25(\text{千克})$$

$$Q_3=0.75\times\text{第 8 项}+0.25\times\text{第 9 项}=0.75\times18+0.25\times18.2=18.05(\text{千克})$$

由组距数列确定四分位数，必须先向上或向下累计次数，根据累计次数确定四分位数的位置，最后确定相关的四分位数。

四分位数不仅利用了中位数的基本方法，而且更加清晰地反映了被研究总体分布的集中情况。所以，四分位数也是描述集中趋势的重要指标。

与四分位数类似，还可以推算十分位数、百分位数。

在实际工作中，国家劳动部门采用“四分位法”测算工资指导线水平，即取下四分位值与上四分位值之间的幅度作为工资合理增长区间。政府运用工资指导线，对国有企业及其他各类企业的工资分配进行指导与调控，使企业工资增长符合经济和社会发展的要求，进一步促进生产力的发展。

4.3.4 应用平均指标要注意的问题

1. 平均指标必须应用于同质总体

所谓同质总体，指总体具有同质性，即社会经济现象的各个单位在被平均的标志上具有同类性质。如果各单位在类型上存在根本差别，则依此计算的平均数不仅不能反映事物的本质和内在规律性，还会歪曲事物真相。例如，在研究职工工资收入变化时，如果把外来务工人员收入，或者是把企业高层管理者的收入与职工的工资收入合在一起求“平均”，则平均结果不能反映职工真实的工资收入变化。只有在同质总体基础上计算应用平均指标，才能有真实的社会经济意义。

2. 平均指标应与分配数列相结合

平均指标的重要特征是把总体各单位的数量差异抽象化了，掩盖了各单位的数量差别及其分布情况。所以在使用平均指标了解现象时，不能只看平均数，而要同时看分配数列的次数分布情况。所以，平均指标要与分配数列结合应用。

【实例 4.27】 某旅游公司对员工进行旅游英语知识测验，平均分数为 72.3 分，37 名员工成绩具体情况见表 4-16。

表 4-16 旅游公司员工旅游英语知识测验成绩

旅游英语测验成绩/分	员工人数/人	旅游英语测验成绩/分	员工人数/人
60 以下	10	80～90	5
60～70	5	90～100	5
70～80	12	合计	37



若单看平均数，只泛泛地了解了该旅游公司平均成绩70多分，比较不错。但如果结合分配数列，就会更具体、更深入地反映问题，即还有10人成绩不及格，占37名员工总数的27%。

3. 平均指标要与分组法相结合

平均指标反映了总体各单位某一数量标志值的一般水平，但却掩盖了各组之间的差异。总体各组之间及各组内的差异往往影响总体的特征和分布规律，各组结构变动也会对总体变动产生影响。为了全面认识总体的特征和分布规律，需要将平均指标与统计分组结合起来。

【实例4.28】 两种商品销售价格资料见表4-17。分别计算6月、7月两款家电平均销售价格并进行分析。

表4-17 两种商品销售价格资料

商品名称	6月销售价格/元	6月销售量/台	7月销售价格/元	7月销售量/台
家电Ⅰ	100	60	90	40
家电Ⅱ	200	40	190	60

解：

利用加权算术平均法计算，结果为

$$6\text{月两款家电平均销售价格}=\frac{100\times60+200\times40}{60+40}=140(\text{元})$$

$$7\text{月两款家电平均销售价格}=\frac{90\times40+190\times60}{40+60}=150(\text{元})$$

7月两款家电平均销售价格比6月两款家电平均销售价格上升了10元。但实际上，从按价格分组的资料看，无论是家电Ⅰ还是家电Ⅱ，7月销售价格均比6月低，即实际结果出现了与总平均数相矛盾的结论。之所以出现这种矛盾的结果，是因为两个月份相比，销售量的结构发生了变化。7月时，价格较高的家电Ⅱ销售量所占比例为60%，比6月时高出20%。正是由于这种权重结构的变化，导致了总平均数与实际情况相矛盾情况的出现。因此，在对实际中经济现象进行描述分析时，应将平均指标与分组法结合应用。

4.4 标志变异指标

平均指标包括数值平均数和位置平均数，它们都是现象一般水平的代表值，反映现象分布的集中趋势。但是，仅仅用集中趋势指标描述现象的特征是不够的。

【实例4.29】 在一次知识竞赛中，男、女两参赛代表队成绩(单位：分)如下：

男代表队：51，65，69，75，81，87，94，95，96，97；

女代表队：74，76，78，79，82，82，83，84，86，86。

这两个参赛代表队的平均分数都是81分，但两队成绩的分散程度却不同。女代表队的成绩比较集中、整齐，即变异较小，从而平均数81分的代表性较好；男代表队的成绩

比较分散，参差不齐，变动较大，用平均分 81 分作为代表，其代表性较差。统计平均指标在反映总体分布一般水平的同时，也掩盖了各标志值的差异性。而标志变异指标恰好弥补了这方面的不足。它能够综合反映总体分布中各标志值的差异程度，显示平均指标的代表性好坏，反映现象分布的离散趋势。

4.4.1　标志变异指标的意义和作用

1. 标志变异指标的意义

标志变异指标是反映总体各单位标志值差异程度的综合指标，又称为标志变动度。它是和统计平均指标相联系的一种综合指标。

平均指标将总体各单位标志值的差异抽象化，从一个侧面反映总体各单位标志值的集中趋势和程度。标志变异指标则从另一个侧面反映总体各单位标志值的差别大小、变动范围和离散程度。平均指标与标志变异指标分别反映同一总体在数量上的集中趋势与离散趋势，两者相辅相成，有助于科学全面地描述客观现象特征，反映现象总体的数量规律。

2. 标志变异指标的作用

(1) 标志变异指标能够反映平均数代表性的大小。平均指标作为总体各单位标志值一般水平的代表性指标，其代表性的大小与标志变异指标大小成反比。即标志变异指标越大，平均数的代表性越小；标志变异指标越小，平均数的代表性越大。

(2) 标志变异指标可以说明现象变动的均衡性、稳定性和节奏性。标志变异指标越大，说明标志值之间的差异程度越大，反映为总体均衡性、稳定性差，节奏感不强；标志变异指标越小，说明标志值之间的差异程度越小，反映为总体均衡性、稳定性好，节奏感强。例如，企业为进行产品质量控制和实施全面质量管理，在考核生产计划执行情况时，除了计算平均完成程度外，还研究计划执行的均衡性和节奏性。已知某公司所属两家企业 2012 年各季度计划完成情况资料见表 4－18。

表 4－18　某公司 2012 年各季度计划完成情况资料

公司所属企业	每季度完成全年计划程度				合　计
	第一季度	第二季度	第三季度	第四季度	
甲企业	24%	25%	25%	26%	100%
乙企业	20%	15%	25%	40%	100%

虽然两家企业均完成全年计划任务，但甲企业各季度完成全年计划程度数值变动不大，说明其标志变动度小，则甲企业工作的均衡性和节奏性好；而乙企业各季度完成全年计划程度数值变动大，高为 40%，低仅有 15%，显示其标志变异指标大，所以，乙企业工作的均衡性和节奏性差。又如，在商品质量检测中，常用标志变异指标说明商品质量的稳定性。

(3) 标志变异指标可以用来确定统计推断的准确程度。在抽样调查中，利用样本指标



推断总体指标时，标志变异指标数值的大小是确定统计推断准确程度，以及计算误差大小不可缺少的重要资料，具体内容将在第7章详述。

4.4.2 标志变异指标的计算及应用

常用的标志变异指标有全距(range)、四分位数间距(interquartile range)、平均差(mean absolute deviation)以及标准差(standard deviation)、标志变异系数(coefficient variation)等。

1. 全距

全距是指总体各单位标志值中最大数值与最小数值之差，又称极差，用“R”表示。它表明各个数值之间最大可能的差距，是一种测量标志变动度的最简单的方法。

对于未分组资料或单项数列资料，R=最大标志值−最小标志值；对于组距数列资料，$R\approx$最高组的上限−最低组的下限。

【实例4.30】 以实例4.29中男、女两参赛代表队成绩资料为例，求全距。

解：

$$男代表队全距：R=97-51=46(分)$$
$$女代表队全距：R=86-74=12(分)$$

男女两参赛代表队平均成绩都是81分，但从全距看，男代表队$R=46$分，表明成绩的差异大，所以平均数的代表性差；女代表队$R=12$分，表明成绩的差异程度较小，所以用81分代表女队成绩，平均数的代表性较好。

全距主要用来说明总体各单位标志值变动总范围。全距越大，各标志值变动范围越大，平均数代表性越差；反之，全距越小，各标志值变动范围越小，平均数代表性越好；全距为0时，平均数的代表性最好。

全距的优点是计算方法简单，意义明确。全距经常用于检查产品质量的稳定性或进行产品质量控制。世界一些国家在提供证券市场行情时，广泛应用最高价、最低价和全距。

全距的缺点是易受极端数值的影响，它只能反映最大值和最小值之间的差距，而不能反映其内部各项数值的差异状况。当标志值有异常值存在时，会直接影响全距大小，使得全距不能充分反映总体分布的离散趋势。全距是一个较粗糙的测定标志变动度的指标。

2. 四分位数间距

四分位数间距是将总体各单位标志值按一定顺序排列后，第三个四分位数与第一个四分位数之差，用“IQR”表示，即$IQR=Q_3-Q_1$。第三个四分位数是处于中位数和最大数值中间的数值，第二个四分位数即中位数，第一个四分位数是处于中位数和最小数值中间的数值。

【实例4.31】 2012年我国对9个主要国家和地区货物出口额资料见表4-19，求四分位数间距。

表 4-19　2012 年我国对主要国家和地区货物出口额

序号(按出口额由小到大)	国家和地区	出口额/亿美元
9	美国	3 518
8	欧盟	3 340
7	中国香港	3 235
6	东盟	2 043
5	日本	1 516
4	韩国	877
3	印度	477
2	俄罗斯	441
1	中国台湾	368

（资料来源：中华人民共和国 2012 年国民经济和社会发展统计公报.）

解：

第一个四分位数的位置是 $\frac{1}{4}\times(9+1)=2.5$，所以，$Q_1=459$ 亿美元；

第三个四分位数的位置是 $\frac{3}{4}\times(9+1)=7.5$，所以，$Q_3=3\ 287.5$ 亿美元。

则四分位数间距 $IQR=Q_3-Q_1=3\ 287.5-459=2\ 828.5$(亿美元)。这表示，2012 年我国前 9 位出口贸易伙伴中一半出口额相差跨度不超过 2 828.5 亿美元。

四分位数间距先将数列一分为四，并略去最大与最小那一组的数值，然后计算余下数值的全距。四分位数间距所包含的数值项数刚好是全部项数的一半。

四分位数间距的优点是克服了全距受极端数值影响的缺点，计算简便。在实际工作中，当人们对极端数值不感兴趣，而对处于当中部分的单位特征更在意时，常使用四分位数间距计算。例如，在房地产行业中，用四分位数间距描述购房者中间的一半人，以及他们所购房屋价格的跨度。但是四分位数间距仍然不能反映总体内部各项标志值的差异状况，所以，它也是一个较粗糙的测定标志变动度的指标。

3. 平均差

全距与四分位数间距的大小都取决于数据中的两个数值。只有能够测度全部标志值对中心位置的偏差，才能对总体的离散程度做出最综合的说明。而要测度全部标志值偏离集中趋势的程度，很自然的设想就是计算各标志值偏离平均数的平均距离。由于各标志值对算术平均数的离差和为零，因此，最简单的做法就是将离差加上绝对值，然后求算术平均数。

平均差是总体各单位标志值与其算术平均数离差绝对值的算术平均数，用符号“MAD”表示。平均差表示总体分布中，各标志值对算术平均数的平均距离。平均差越大，说明各标志值差异程度越大，平均数代表性越差；平均差越小，说明各标志值差异程度越小，平均数代表性越好。



根据所掌握资料不同，平均差可分为简单平均差和加权平均差两种。其基本公式为

$$\text{简单平均差：} MAD = \frac{\sum |x-\bar{x}|}{n} \text{（根据未分组资料计算时）} \quad (4\text{-}24)$$

$$\text{加权平均差：} MAD = \frac{\sum |x-\bar{x}| f}{\sum f} \text{（根据分组资料计算时）} \quad (4\text{-}25)$$

【实例 4.32】 一家生产公司新引进一条产品生产线生产计算机，投产后前5个星期产量资料分别为5台、9台、16台、17台、18台，求平均差。

解：

首先计算平均数：

$$\bar{x} = \frac{\sum x}{n} = \frac{5+9+16+17+18}{5} = 13 \text{（台）}$$

然后计算平均差：

$$MAD = \frac{\sum |x-\bar{x}|}{n} = \frac{|5-13|+|9-13|+|16-13|+|17-13|+|18-13|}{5} = 4.8 \text{（台）}$$

【实例 4.33】 对某地区90家电子商务公司从业人员进行调查，员工人数资料见表4-20，求平均差。

表4-20 90家电子商务公司员工人数资料及计算

员工人数/人	公司数(f)/家	组中值(x)	xf	$\|x-\bar{x}\|$	$\|x-\bar{x}\|f$
20以下	32	10	320	32.89	1 052.48
20～40	16	30	480	12.89	206.24
40～60	13	50	650	7.11	92.43
60～80	10	70	700	27.11	271.1
80～100	19	90	1710	47.11	895.09
合计	90	—	3 860	—	2 517.34

解：

首先计算平均数：

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{3\,860}{90} \approx 42.89 \text{（人）}$$

然后计算平均差：

$$MAD = \frac{\sum |x-\bar{x}| f}{\sum f} = \frac{2\,517.34}{90} \approx 27.97 \text{（人）}$$

平均差是根据全部标志值计算出来的，反映了每个标志值与平均数的平均离散程度。与全距相比，受极端数值的影响较小，是比全距更优良的标志变异指标。在统计实践中，一般在产品质量控制分析中应用平均差。

但计算平均差时，须对离差取绝对值，这为平均差的代数运算带来了许多不便，从而使它的应用受到了限制。平均差并不是测定总体离散趋势的最好方法，在实际中，最常用的标志变异指标是标准差。

4. 标准差

标准差是总体各单位标志值与其算术平均数离差平方的算术平均数的平方根。又称为均方差，用“σ”表示。标准差的平方称为方差(variance)，用“σ^2”表示。标准差是标志变异指标中最重要，也是最常用的指标。

标准差的实质与平均差基本相同，也表示各个标志值与平均数的平均离散程度。由于标准差采用平方而不是加绝对值的方法来消除各标志值与算术平均数离差的正负号，因此它的应用比平均差更广泛。

根据所掌握资料不同，标准差可分为简单标准差和加权标准差两种。其基本公式为

$$\text{简单标准差：}\sigma=\sqrt{\frac{\sum(x-\bar{x})^2}{n}}\text{（根据未分组资料计算时）} \tag{4-26}$$

$$\text{加权标准差：}\sigma=\sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}\text{（根据分组资料计算时）} \tag{4-27}$$

【实例 4.34】 根据实例 4.32 资料计算标准差。

解：

(1) 计算算术平均数：

$$\bar{x}=\frac{\sum x}{n}=\frac{5+9+16+17+18}{5}=13\text{(台)}$$

(2) 计算各单位标志值与其算术平均数的离差：

“$x-\bar{x}$”分别为−8、−4、3、4 和 5。

(3) 计算各单位标志值与算术平均数的离差平方和：

$$\left(\sum x-\bar{x}\right)^2=64+16+9+16+25=130$$

(4) 计算离差平方和的算术平均数：

$$\frac{\sum(x-\bar{x})^2}{n}=\frac{130}{5}=26\text{(台)}$$

(5) 开平方，得标准差：

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2}{n}}=\sqrt{26}\approx 5.1\text{(台)}$$

实例 4.34 是资料未分组情况下简单标准差的计算。当资料分组时，算术平均数与标准差的计算均须采用加权平均法。加权平均法计算标准差的步骤与简单平均法基本相同。

【实例 4.35】 某地区 90 家电子商务公司员工人数资料见表 4－21，计算电子商务公司员工人数标准差。



表 4-21 90家电子商务公司员工人数标准差计算

员工人数/人	公司数(f)/家	组中值(x)	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
20以下	32	10	320	−32.89	1 081.752 1	34 616.067 2
20～40	16	30	480	−12.89	166.152 1	2 658.433 6
40～60	13	50	650	7.11	50.552 1	657.177 3
60～80	10	70	700	27.11	734.952 1	7 349.521
80～100	19	90	1 710	47.11	2 219.352 1	42 167.689 9
合计	90	—	3 860	—	—	87 448.889

解：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{3\ 860}{90}\approx 42.89(\text{人})$$

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}=\sqrt{\frac{87\ 448.889}{90}}=\sqrt{971.654\ 322\ 2}\approx 31.17(\text{人})$$

因此，这90家电子商务公司员工人数标准差约为31.17人。

标准差是根据全部标志值计算的，任何一个标志值发生变化，都会使标准差发生变化。所以标准差反应灵敏，能准确反映总体分布的离散趋势。另外，它克服了平均差计算中用绝对值消除离差正负号带来的运算问题，在计算上比平均差更简便。标准差是实际应用最广泛的标志变异指标。在统计分析中，计算算术平均数的同时，往往同时计算标准差。在相关、回归、抽样推断、统计预测等方面均要应用标准差。

但是由于标准差的大小与平均数有关，且标准差带有与标志值相同的计量单位，因此标准差不适宜用于比较平均数相差较大或计量单位不同的两总体分布的离散程度。

5. 标志变异系数

标志变异系数又称离散系数，是用标志变异指标与相应的算术平均数对比，反映总体各单位标志值之间离散程度的指标。它是以相对数表示的标志变异指标。

前面介绍的各种标志变异指标都是表示标志变动程度的绝对指标，其数值的大小，还要受总体单位标志值本身水平高低的影响。在分析比较两组数字资料的离散趋势时，如果两组数字的平均数相差较大，或数字的计量单位不同，则不能使用绝对变异指标，必须进一步计算标志变异系数。标志变异系数包括全距系数、平均差系数、标准差系数等，实际统计分析工作中最常用的是标准差系数。

标准差系数的计算公式为

$$CV=\frac{\sigma}{\bar{x}}\times 100\% \tag{4-28}$$

式中：CV 为标准差系数。

【实例 4.36】 某学院市场营销专业2012级有两个班，英语期末考试平均成绩分别为 $\overline{x_1}=86$ 分，$\overline{x_2}=75$ 分；标准差分别为 $\sigma_1=12$ 分，$\sigma_2=11$ 分。计算标准差系数，并分析哪个班

的成绩比较均匀，哪个班的平均成绩更具有代表性。

解：

$$\text{一班成绩的标准差系数：} CV_1 = \frac{\sigma}{\bar{x}} \times 100\% = \frac{12}{86} \times 100\% \approx 13.95\%$$

$$\text{二班成绩的标准差系数：} CV_2 = \frac{\sigma}{\bar{x}} \times 100\% = \frac{11}{75} \times 100\% \approx 14.67\%$$

虽然一班成绩的标准差大于二班成绩的标准差，但由于两班平均成绩相差较大，不能由此说明一班平均成绩代表性比二班的差。为了说明两班考试成绩的离散程度，可以比较两班的标准差系数。由于一班成绩的标准差系数13.95%小于二班成绩的标准差系数14.67%，所以得出结论：一班的考试成绩比较均匀，一班平均成绩更有代表性。

标准差系数是无名数，在应用时不受计量单位和标志值水平的限制，消除了不同总体之间在计量单位、平均水平方面的不可比性。适用于对比分析平均水平不同或计量单位不同的两组数据的离散程度的大小。标准差系数大的，说明数据的离散程度大；标准差系数小的，说明数据的离散程度小。

总之，反映总体离散趋势的各种标志变异指标，适用于不同的总体状况。其中，标准差应用最广。当需要对不同总体数据的离散程度进行比较时，标准差系数最常用。

本 章 小 结

本章共有4节，安排4项内容。一是总量指标，二是相对指标，三是平均指标，四是标志变异指标。

总量指标又称绝对数，用来说明客观现象在一定时间、地点、条件下的总规模或总水平，是认识社会经济现象的起点，也是计算相对指标和平均指标的基础。总量指标根据反映的内容不同，分为总体单位总量与总体标志总量。根据反映的时间状况不同，分为时期指标与时点指标。总量指标在应用时，应注意正确确定指标的内涵及不同指标的界限，注意现象的同类性，不同类现象不能相加汇总，计量单位必须统一，在实际应用时总量指标一般不单独使用。

相对指标又称为相对数，它是由两个相互联系的指标对比来表示相关现象之间联系程度的指标。常用的相对指标有结构相对指标、比例相对指标、比较相对指标、强度相对指标、动态相对指标和计划完成程度相对指标。运用这些相对指标对社会经济现象进行比较、评价时，必须注意对比指标的可比性。此外，还必须注意两点：一是正确选择对比的基数；二是相对指标与总量指标结合应用，多种相对指标要结合应用。

平均指标是反映现象总体各单位某一数量标志值的典型水平、一般水平和代表性水平。它除了反映总体各单位分布的集中趋势外，还可以对现象在不同空间、时间上进行比较分析。此外，进行现象之间的依存关系分析和数量上的估算时离不开平均数，评价事物好坏常以平均数作为参考依据。平均指标有动态和静态平均数之分，本章讨论的是静态平均数，主要包括数值平均数和位置平均数。

数值平均数是根据总体各单位标志值计算的平均数，包括算术平均数、调和平均数、几何平均数。算术平均数是实际中最常用、最基本的平均指标。统计中平均指标一般指算



术平均数。在经济统计中，经常因为无法直接得到被平均标志值的相应次数资料而采用调和平均数的形式来计算，调和平均数可看作算术平均数的变形。当数据资料是比率变量时，采用几何平均法进行计算。数值平均数易受极端数值的影响，只适用于对数值型数据分布集中趋势的反映。

位置平均数是根据总体各单位标志值所处的位置确定的，主要包括众数、中位数和四分位数。数值平均数和位置平均数所包含的各指标各有特点和适用条件，计算方法不同，对同一资料计算结果也不同。因此，必须根据研究目的和总体分布特征，选择适当的平均指标。

标志变异指标是反映总体各单位标志值差异程度的综合指标。标志变异指标越小，则总体各单位分布的离散程度越小，平均数的代表性越大。经常使用的标志变异指标有全距、四分位数间距、平均差、标准差和标志变异系数。在统计实践中，算术平均数和标准差是描述统计特征时最常用的、反映集中趋势和离散趋势的两个最重要指标。



应用与拓展

2012 年城镇非私营单位就业人员年平均工资 46 769 元

2013 年 5 月 17 日，国家统计局发布的数据显示，2012 年全国城镇非私营单位就业人员年平均工资为 46 769 元，与 2011 年的 41 799 元相比，增加了 4 970 元，同比名义增长 11.9%，增幅下降 2.5 个百分点。其中，在岗职工平均工资 47 593 元，同比名义增长 12.1%，增幅下降 2.2 个百分点。扣除物价因素，2012 年全国城镇非私营单位就业人员年平均工资实际增长 9.0%。

如表 4-22 所示，分四大区域看，2012 年城镇非私营单位就业人员年平均工资由高到低排列是东部、西部、中部和东北，分别为 53 444 元、41 959 元、40 110 元和 39 245 元；同比名义增长率从高到低依次为西部 13.1%、东北 12.8%、中部 12.3%和东部 10.9%。

表 4-22　2012 年城镇非私营单位分地区就业人员年平均工资

地　　区	2011 年/元	2012 年/元	名义增长率/%
合计	41 799	46 769	11.9
东部	48 183	53 444	10.9
中部	35 718	40 110	12.3
西部	37 091	41 959	13.1
东北	34 784	39 245	12.8

如表 4-23 所示，分行业门类看，年平均工资最高的三个行业分别是金融业 89 743 元，是全国平均水平的 1.92 倍；信息传输、软件和信息技术服务业 80 510 元，是全国平均水平的 1.72 倍；科学研究、技术服务业 69 254 元，是全国平均水平的 1.48 倍。年平均工资最低的三个行业分别是农、林、牧、渔业 22 687 元，是全国平均水平的 49%；住宿和餐饮业 31 267 元，是全国平均水平的 67%；水利、环境和公共设施管理业 32 343 元，是全国平均水平的 69%。最高与最低行业平均工资之比是 3.96∶1，比 2011 年的 4.17∶1 差距有所缩小。

表 4-23　2012 年城镇非私营单位就业人员分行业年平均工资

行　　业	2011 年/元	2012 年/元	名义增长率/%
合计	41 799	46 769	11.9
农、林、牧、渔业	19 469	22 687	16.5
采矿业	52 230	56 946	9.0
制造业	36 665	41 650	13.6
电力、热力、燃气及水生产和供应业	52 723	58 202	10.4
建筑业	32 103	36 483	13.6
批发和零售业	40 654	46 340	14.0
交通运输、仓储和邮政业	47 078	53 391	13.4
住宿和餐饮业	27 486	31 267	13.8
信息传输、软件和信息技术服务业	70 918	80 510	13.5
金融业	81 109	89 743	10.6
房地产业	42 837	46 764	9.2
租赁和商务服务业	46 976	53 162	13.2
科学研究、技术服务业	64 252	69 254	7.8
水利、环境和公共设施管理业	28 868	32 343	12.0
居民服务、修理和其他服务业	33 169	35 135	5.9
教育	43 194	47 734	10.5
卫生和社会工作	46 206	52 564	13.8
文化、体育和娱乐业	47 878	53 558	11.9
公共管理、社会保障和社会组织	42 062	46 074	9.5

如表 4-24 所示，从登记注册类型看，股份有限公司的年平均工资最高，为 56 254 元，是全国平均水平的 1.20 倍；其次为外商投资企业，为 55 888 元，是全国平均水平的 1.20 倍；第三位是国有单位，为 48 357 元，是全国平均水平的 1.03 倍。年平均工资最低的为集体单位，为 33 784 元，是全国平均水平的 72%。

表 4-24　2012 年城镇非私营单位分登记注册类型就业人员年平均工资

登记注册类型	2011 年/元	2012 年/元	名义增长率/%
合计	41 799	46 769	11.9
国有	43 483	48 357	11.2
集体	28 791	33 784	17.3
股份合作	36 740	43 433	18.2



续表

登记注册类型	2011年/元	2012年/元	名义增长率/%
联营	36 142	42 083	16.4
有限责任公司	37 611	41 860	11.3
股份有限公司	49 978	56 254	12.6
其他内资	29 961	34 694	15.8
港澳台商投资	38 341	44 103	15.0
外商投资	48 869	55 888	14.4

（资料来源：人民网，2013-05-17.）



习题与实训

一、单项选择题

1. 年末人口数、出生人口数两指标(　　)。

A. 前者是时期指标，后者是时点指标

B. 前者是时点指标，后者是时期指标

C. 两者都是时期指标

D. 两者都是时点指标

2. 某旅游企业5月接待三批旅游人员，第一批接待旅游人数1 000人，过夜4天；第二批接待800人，过夜5天；第三批接待1 200人，过夜5天。则该旅游企业5月服务工作总量为(　　)。

A. 10 000人天数　　B. 14 000人天数

C. 12 000人天数　　D. 16 000人天数

3. 某地区年末面积1万平方千米，年末总人口3 420万人，其中男性人口1 745万人，女性人口1 675万人，总人口比上年末净增24万人，年内新出生人口51万人。则利用该数据可计算的相对指标有(　　)。

A. 比例相对指标、动态相对指标、强度相对指标、计划完成程度相对指标

B. 结构相对指标、动态相对指标、强度相对指标

C. 结构相对指标、比例相对指标、动态相对指标

D. 结构相对指标、比例相对指标、动态相对指标、强度相对指标

4. 华星企业生产的变速自行车上年实际成本为450元，本年计划降低4%，实际降低了3.5%，则成本降低计划完成程度为(　　)。

A. 100.52%　　B. 99.52%　　C. 87.5%　　D. 99.48%

5. 按照计划，现年产量比上年应增加30%，实际比计划少完成10%，同上年相比现年产量的实际增长程度为(　　)。

A. 10%　　B. 17%　　C. 20%　　D. 40%

6. 下面关于平均指标的描述错误的是(　　)。

A. 可以反映总体各单位分布的集中趋势

B. 它是一个抽象化的数值

C. 它是某一个单位的具体数值

D. 可以作为评价事物的参考依据

7. 对数值平均数描述错误的是(　　)。

A. 按照计算方法不同，分为算术平均数、调和平均数、几何平均数

B. 是静态平均数的一种

C. 都是根据总体各单位标志值计算出来的

D. 常见的主要有众数、中位数和四分位数等

8. 随机调查某城市 100 户家庭，得到家庭订阅报纸杂志份数资料见表 4-25。

表 4-25　家庭订阅报纸杂志份数资料

报纸杂志数/份	0	1	2	3	4	5	合计
家庭数/户	9	54	21	12	2	2	100

则根据这份资料计算的众数是(　　)。

A. 21　　B. 1　　C. 54　　D. 2

9. 在甲乙两个变量数列中，若甲数列标准差小于乙数列标准差，则两个变量数列平均数代表性相比较(　　)。

A. 甲数列的平均数代表性高于乙数列

B. 两个数列的平均数代表性相同

C. 乙数列的平均数代表性高于甲数列

D. 不能确定

10. 两个总体平均数不等，但标准差相等，则(　　)。

A. 两个平均数代表性相同　　B. 平均数大代表性小

C. 平均数小代表性小　　D. 无法进行正确判断

二、多项选择题

1. 下列属于时点指标的有(　　)。

A. 某地区人口数　　B. 某地区人口死亡数

C. 某市在校学生数　　D. 某地区基本建设投资额

E. 某农场每年拖拉机台数

2. 下列统计指标中属于强度相对指标的是(　　)。

A. 人口密度　　B. 人均国民收入

C. 人口死亡率　　D. 城镇居民人均可支配收入

E. 经济发展速度

3. 某厂计划现年的能源消耗比上年降低 5%，产值增长 8%。实际的完成情况是能源消耗降低了 6%，产值上升了 7%。则(　　)。



A. 能源消耗的计划完成程度是98.95%

B. 能源消耗的计划完成程度是110%

C. 产值的计划完成程度是99.07%

D. 能源消耗完成了计划而产值没有完成计划

E. 产值完成了计划而能源消耗没有完成计划

4. “2012年我国全年农村居民人均纯收入7 917元，比上年增长13.5%，扣除价格因素，实际增长10.7%；农村居民人均纯收入中位数为7 019元，增长13.3%。城镇居民人均可支配收入24 565元，比上年增长12.6%，扣除价格因素，实际增长9.6%；城镇居民人均可支配收入中位数为21 986元，增长15.0%。农村居民食品消费支出占消费总支出的比例为39.3%，城镇为36.2%。”在这段文字中(　　)。

A. 有相对指标但没有总量指标

B. 既有强度相对指标又有结构相对指标、动态相对指标

C. 既有平均指标又有相对指标

D. 没有比较相对指标

E. 既有相对指标又有总量指标

5. 两种价格不同的食品，在销售量都增加的情况下，两种食品的平均价格(　　)。

A. 肯定变动

B. 肯定不变

C. 可能变动也可能不变

D. 若销售量成比例增加，则平均价格不变

E. 若销售量增加数相同，则平均价格不变

6. 马克思曾明确指出“如果1个工人每星期的工资是2先令，后来他的工资提高到4先令，那么工资水平就提高了100%……所以不应当为工资水平提高的动听的百分比所迷惑。我们必须经常这样问：原来的工资数是多少?”这段话提示我们(　　)。

A. 数量指标要与质量指标结合应用

B. 相对指标要与总量指标结合应用

C. 各种相对指标要结合应用

D. 相对指标要与平均指标结合应用

E. 计算相对指标时应保证两个对比指标的可比性

7. 下列平均数要用几何平均法计算的有(　　)。

A. 生产同种产品的3个车间的平均合格率

B. 前后工序的3个车间的平均合格率

C. 以复利支付利息的年平均利率

D. 平均发展速度

E. 平均劳动生产率

8. 在各种平均指标中，不受极端值影响的平均指标是(　　)。

A. 算术平均数

B. 几何平均数

C. 调和平均数

D. 众数

E. 中位数

9. 中位数是(　　)。

A. 由标志值在数列中的位置决定的

B. 根据标志值出现的次数决定的

C. 总体一般水平的代表值

D. 不受总体中极端数值的影响

E. 总体各单位的平均指标

10. 某人购房欲贷款12万元，根据其资信水平，贷款10万元的年利率是8%，若增加2万元，则这12万元贷款的年利率变为10%，则(　　)。

A. 增加的2万元贷款的年利率是20%

B. 增加的2万元贷款的年利率是15%

C. 增加的2万元贷款的年利率是12%

D. 计算增加的2万元贷款的年利率利用的是加权调和平均法

E. 计算增加的2万元贷款的年利率利用的是加权算术平均法

三、判断题

1. 2012年我国出生人口性别比为117∶100，这属于比较相对指标。　　(　　)

2. 城镇居民恩格尔系数和农村居民恩格尔系数都是结构相对指标，它们之和应为1。
(　　)

3. 2011年北京市人均国内生产总值为79 265元/人，上海为81 772元/人，则上海是北京的103.16%，北京是上海的96.93%。　　(　　)

4. 企业计划规定，2012年第四季度的单位产品成本比去年同期降低2%，实际执行结果降低了1%，则企业仅完成单位产品成本计划的一半。　　(　　)

5. 数值平均数属于静态平均数，而位置平均数属于动态平均数。　　(　　)

6. 众数是总体中出现最多的次数。　　(　　)

7. 在对比两个不同总体各单位标志值的离散程度时，应采用标准差而不是全距或平均差指标进行比较。　　(　　)

8. 标准差系数越大，说明总体中各单位标志值的变异程度越大，则平均指标的代表性就越好。　　(　　)

9. 某组学生身高(单位：厘米)分别为157、159、162、163、164、167、168、172，则中位数是163.5。　　(　　)

10. 加权算术平均数计算公式为$\bar{x}=\frac{\sum xf}{\sum f}$，由此可看出，影响$\bar{x}$大小的因素有两个，一是$x$，另一个是$f$。　　(　　)

四、填空题

1. “2006年，香港的就业人数达到346万人，比1996年增加了39万人，增长12.7%；2006年，香港幼儿园、小学和中学教育机构分别达到1 015所、668所和565所，分别比1996年增长38.3%、下降22.0%和增长12.3%。”在这段文字中，有两种指标在结合运用，它们是________和________。



2. 某地区某年的财政总收入 1 248.5 亿元是总量指标，从反映总体的时间上看，该指标是________指标；从反映总体的内容上看，该指标是________指标。

3. 某市 2012 年单位 GDP 电耗(单位：千瓦时/万元)计划下降 4%，实际下降了 4.84%，则该市超额完成计划________，实际比计划多下降了________个百分点。

4. 在报考公务员的 15 982 人中，有 198 人被录取，则录取率为________，这是________相对指标。

5. 平均指标反映总体数据的________趋势；标志变异指标反映总体数据的________程度。

6. 某鱼贩按 15 元/千克的均价购得一批鱼，进行分装销售时，发现大鱼占 40%，小鱼占 60%，若小鱼按 12 元/千克出售，则大鱼至少售价为________才能不亏本。

7. 已知 3 种产品的合格率分别为 58%、76%、91%，则 3 种产品的平均合格率为________。

8. 已知某产品流水线上 3 道工序产品合格率分别为 94%、91%、97%，则该产品平均合格率为______。

9. 现有一数列：3、9、27、81、243、729、2187，则反映其平均水平最好采用______指标。

10. 将某邮局外发邮包样本的重量近似到盎司为 21、18、30、12、14、17、28、10、16、25。则这组数据的均值是________，中位数是________，众数是________，极差是________，四分位数间距是________。

五、应用能力训练题

1. 某公司在年终统计分析报告中写道：我公司今年销售收入计划规定 2 500 万元，实际完成了 2 550 万元，超额完成计划 2%；销售利润率计划规定 8%，实际为 12%，超额完成计划 4%；劳动生产率计划规定比去年提高 5%，实际比去年提高 5.5%，完成计划 110%；产品单位成本计划规定比去年下降 3%，实际比去年下降 2.5%，实际比计划多下降 0.5 个百分点。

请指出上述分析报告中哪些指标计算有错误，将其改正过来。

2. 某市某五年计划规定，计划期最末一年甲产品产量应达到 50 万吨的水平，实际情况见表 4-26。

表 4-26　某市五年计划的实际情况

时　间	第一年	第二年	第三年	第四年				第五年			
				第一季度	第二季度	第三季度	第四季度	第一季度	第二季度	第三季度	第四季度
产量/万吨	42	45	47	10	11	13	13	13	12	14	13

试计算该市甲产品产量五年计划完成程度和提前完成计划的时间。

3. 某公司 2012 年计划产值为 3 160 万元，结果计划完成了 110%，2012 年产值计划比 2011 年增长 8.5%，试计算实际产值 2012 年比 2011 年的增长率。

4. 某生产公司 2009 年经过战略调整，改革措施到位，公司取得了辉煌的业绩。在年终总结报告中，公司管理人员生动、丰富又有说服力地表达了生产业绩的“好”。下面是部分摘录内容。

2009年我公司实现产量110万台，这是公司历史上最高产量。与我们年初定的计划90万台相比，我们超额22.2%完成了任务，多生产了20万台；与我们去年产量相比，今年我们再上一个新台阶，比去年年产85万台的产量多29.4%，比去年多生产了25万台。这是一个骄人的成绩，因为我们同行业的竞争对手A厂最高水平是108万台，我们已经超出1.85%，还多出2万台。这个成绩给了我们很强的信心，我们"十一五"计划总产量是400万台，仅2009年一年完成的产量，就占五年计划的27.5%，累计到2009年年末，我们已经生产340万台，完成"十一五"计划的85%，提前完成五年计划已成定局。但我们在成绩面前要保持清醒，我们公司非生产工人人数较多，所以，全员劳动生产率指标还不算太高，我们应该奋发进取，不断创新，争取更大成绩！

写出上述文字中出现的所有指标，它们分别属于哪一种总量指标？哪一种相对指标？通过此例，拓展思路，学会在实际应用时纵横对比，多角度、全方位描述现象。

5. 表4-27是美国《财富》杂志2012年7月9日公布的2012年度世界500强企业中前20强企业的营业收入资料，请求出营业收入的平均数和中位数。

表4-27　2012年度世界500强企业中前20强企业的营业收入资料

排名	公司名称(中英文)	营业收入/百万美元	国　家
1	荷兰皇家壳牌石油公司(ROYAL DUTCH SHELL)	484 489.0	荷兰
2	埃克森美孚(EXXON MOBIL)	452 926.0	美国
3	沃尔玛公司(WAL-MART STORES)	446 950.0	美国
4	英国石油公司(BP)	386 463.0	英国
5	中国石油化工集团公司(SINOPEC GROUP)	375 214.0	中国
6	中国石油天然气集团公司(CHINA NATIONAL PETROLEUM)	352 338.0	中国
7	国家电网公司(STATE GRID)	259 141.8	中国
8	雪佛龙股份有限公司(CHEVRON)	245 621.0	美国
9	康菲石油公司(CONOCOPHILLIPS)	237 272.0	美国
10	丰田汽车公司(TOYOTA MOTOR)	235 364.0	日本
11	道达尔公司(TOTAL)	231 579.8	法国
12	大众公司(VOLKSWAGEN)	221 550.5	德国
13	日本邮政控股公司(JAPAN POST HOLDINGS)	211 018.9	日本
14	嘉能可国际公司(GLENCORE INTERNATIONAL)	186 152.0	瑞士
15	俄罗斯天然气工业股份公司(GAZPROM)	157 830.5	俄罗斯
16	意昂集团(E. ON)	157 057.1	德国
17	埃尼石油公司(ENI)	153 675.5	意大利
18	荷兰国际集团(ING GROUP)	150 570.7	荷兰
19	通用汽车公司(GENERAL MOTORS)	150 276.0	美国
20	三星电子(SAMSUNG ELECTRONICS)	148 944.4	韩国



6. 某集团下属 20 家公司在 2013 年第一季度产值计划完成情况见表 4-28。

表 4-28 20 家公司产值计划完成情况

计划完成程度	公司数/家	计划任务数/个	计划完成程度	公司数/家	计划任务数/个
90%～100%	3	80	110%～120%	5	120
100%～110%	12	400	合计	20	

请计算平均计划完成程度。

7. 某投资者在 3 个不同时间分别以每股 6 元、5 元和 4 元各购买了 6 万元的甲股票，如果不计交易成本，则该投资者拥有甲股票的每股平均持仓成本为多少？

8. 投资银行某笔投资的年利率是按复利计算的，过去 25 年的年利率资料见表 4-29。

表 4-29 过去 25 年的年利率资料

年利率	年　　数	年利率	年　　数	年利率	年　　数
3%	1	8%	8	13%	2
4%	4	10%	10	合计	25

则这 25 年的平均年利率为多少？

9. 中国互联网络中心(CNNIC)2014 年 1 月发布第 33 次中国互联网络发展状况统计报告显示，中国互联网用户年龄分布资料见表 4-30。

表 4-30 中国互联网用户年龄分布资料

年龄/岁	10 以下	10～19	20～29	30～39	40～49	50～59	60 及以上
比例	1.9%	24.1%	31.2%	23.9%	12.1%	5.1%	1.9%

请计算中国互联网用户年龄的平均数、标准差、众数、中位数。

10. 一种儿童玩具需要人工组装，现有 3 种可供选择的组装方法。为检验哪种方法更好，随机抽取 15 名工人，让他们分别用 3 种方法组装。15 名工人分别用 3 种方法在相同的时间内组装的产品数量见表 4-31。

表 4-31 工人用 3 种方法组装的产品数量　　单位：个

第一种方法	127	129	129	130	130	130	129	131	127	128	128	128	128	132	125
第二种方法	162	164	165	167	168	164	165	168	170	163	164	166	165	166	167
第三种方法	162	168	169	169	170	152	150	169	174	170	161	169	160	171	167

请运用本章所学知识，说明采用什么指标评价组装方法的好坏。如果让您选择一种方法，您将选择第几种方法？请计算并说明理由。

11. 对10名成年人和10名幼儿的身高进行抽样调查，得到资料见表4-32。

表4-32　幼儿组及成年组身高资料　　单位：厘米

幼儿组身高	74	70	68	71	69	75	72	68	73	73
成年组身高	168	166	169	180	177	172	170	172	174	173

试比较：成年组和幼儿组哪一组的身高差异大。

12. 在股票市场上，高收益率往往伴随着高风险。表4-33是某交易所3只股票连续10个交易日的收盘价格。

(1) 您认为应用什么统计指标来反映投资风险。

(2) 分别计算3只股票价格的标准差、平均数及标准差系数。

(3) 如果您进行股票投资，您会选哪只股票。

表4-33　某交易所3只股票连续10个交易日的收盘价格

交易日 / 股票收盘价	1	2	3	4	5	6	7	8	9	10
甲股票收盘价/元	22	21	26	28	33	24	30	33	28	25
乙股票收盘价/元	23	24	25	25	29	28	26	29	30	31
丙股票收盘价/元	22	18	17	24	30	34	36	33	30	26

第 5 章

时间数列

理论目标

(1) 掌握时间数列的概念和种类。
(2) 掌握编制时间数列的基本原则。
(3) 掌握时间数列的分析方法。

能力目标

(1) 能正确编制时间数列。
(2) 能对时间数列进行水平分析和速度分析。
(3) 能够根据时间数列归纳现象长期趋势和季节变动规律，并进行预测。

做一做

浙江的李先生在上海浦东地区开了一家超市。由于开业前做过市场调研，开业后加强营销管理，超市发展良好，销售规模越来越大。表 5-1 是该超市 2006—2012 年年销售额资料。请根据表中资料计算并回答，2006—2012 年李先生的超市平均年销售额为多少？超市销售额的平均增长速度如何？照此规律发展，2013 年销售额将会达到多少？

表 5-1　超市 2006—2012 年年销售额

年　　份	2006	2007	2008	2009	2010	2011	2012
销售额/百万元	5.6	11.4	17.1	24	31.3	38	43.9

想一想

根据表 5-1 中的资料，我们不能莽然计算 2006—2012 年超市平均年销售额，因为这是一种跨越时间年度的动态平均，与前面章节的静态平均应用条件不一样；速度是日常生活中经常出现的词汇，“速度”到底为何意？“平均速度”如何计算？怎样根据以往的数据，从中总结出规律性的结论，结合目前客观实际，对未来进行预测？这些问题的解决需要用到一个新的统计工具——时间数列，本章将简要介绍有关时间数列的理论和方法。

5.1 时间数列的编制

5.1.1 时间数列的概念

所谓时间数列(time-series)，就是将某一统计指标的数值按时间的先后顺序排列起来所形成的数列，又称为动态数列。例如，表 5-1 所列的超市 2006—2012 年年销售资料就是一个时间数列。

从表 5-1 中可以看出，时间数列一般是由两个基本要素构成的：一是被研究现象所属的时间，如表 5-1 中的 2006 年、2007 年等；二是现象在各个时间上的统计指标数值，如表 5-1 中的销售额。在时间数列中，指标数值也称为发展水平。

时间数列在统计和经济分析中，有着极为重要的作用。当回顾以往工作进行总结时，可以通过对时间数列的编制和分析，描绘社会经济现象变化的过程；当立足现在对现状态势进行分析时，可以运用时间数列，将不同地区或国家进行对比分析；当展望未来进行经济预测时，可以研究时间数列中现象发展变化的方向、速度、水平、趋势和规律性，为社会经济现象的预测提供依据。

5.1.2 时间数列的种类

时间数列按其统计指标表现的不同形式可分为 3 种，即总量指标时间数列、相对指标时间数列和平均指标时间数列。其中总量指标时间数列是最基本的时间数列，相对指标时间数列和平均指标时间数列则是在其基础上派生而成的数列。

1. 总量指标时间数列

总量指标时间数列又称为绝对数时间数列，是由一系列同类总量指标按时间顺序排列形成的时间数列，它反映某种社会经济现象在一段时间内达到的绝对水平及其增减变化情况。例如，表 5-1 就反映了超市在 2006—2012 年各个时期销售额所达到的绝对水平。总量指标时间数列按指标所反映的社会经济现象的时间状况不同，又可分为时期指标数列和时点指标数列两种，它们分别简称为时期数列和时点数列。

1）时期数列及其特点

在总量指标时间数列中，如果指标数值都是反映某种社会经济现象在一定时期内发展过程的总量，则这种时间数列就称为时期数列。表 5-1 所列的数列就是一个时期数列，



其中每项指标分别反映2006—2012年各年的销售额的总量。时期数列的特点如下。

(1) 时期数列的每一指标数值总是和一定的时期相对应的，数列中各指标值可以累加，相加之后表示现象更长一段时期内的发展总量。

(2) 时期数列中每个指标数值的大小与其所属的时期长短有关，时期越长，指标数值越大；时期越短，指标数值越小。

(3) 时期数列中各指标数值通常是通过经常性调查及连续不断登记、汇总而取得的。

2) 时点数列及其特点

在总量指标时间数列中，如果指标数值都是反映某种社会经济现象在某一时刻(时点)上的状况及水平，则这种时间数列就称为时点数列。表5-2所列的数列就是时点数列。

表5-2 2007—2012年年末全国总人口数

年 份	2007	2008	2009	2010	2011	2012
年末全国总人口数/万人	132 129	132 802	133 474	134 100	134 735	135 404

与时期数列相对应，时点数列具有以下特点。

(1) 时点数列中的每一指标只表明社会经济现象在一定时点上的水平，各指标值不能累加，若相加就产生了重复计算，且无实际意义。

(2) 时点数列中每个指标数值的大小与其时间间隔长短无直接关系。

(3) 时点数列中各指标数值一是通过一次性调查登记取得的，即时点数列不具有连续统计的特点。

在运用统计资料进行分析时，正确区分时期数列和时点数列具有十分重要的意义，因为根据这两种数列计算平均发展水平时，其计算方法是截然不同的。

2. 相对指标时间数列

相对指标时间数列又称为相对数时间数列，是由一系列相对指标数值按时间顺序排列形成的时间数列，它反映社会经济现象之间数量对比关系的发展变化过程和趋势。由于相对指标一般表现为两个相关的总量指标之比，因此，两个时期指标、两个时点指标或一个时期指标、一个时点指标的对比，都可以形成相对指标时间数列。例如，表5-3就是一个由“第三产业增加值”与“国内生产总值”这两个时期指标之比所形成的相对指标时间数列。

表5-3 2007—2012年我国第三产业增加值占国内生产总值比重

年 份	2007	2008	2009	2010	2011	2012
第三产业增加值占国内生产总值比重	39.1%	40.1%	42.6%	43.0%	43.1%	44.6%

由于相对指标数值的基数不同，因此，相对指标时间数列中的各项数值不能直接相加。

3. 平均指标时间数列

平均指标时间数列又称为平均数时间数列，是由一系列同类平均指标数值按时间顺序

排列形成的时间数列，它反映社会经济现象一般水平的变化过程和发展趋势。表5-4就是平均指标时间数列。

表5-4 “十一五”期间某公司职工年平均工资

年 份	2006	2007	2008	2009	2010
职工年平均工资/(元/人)	22 130	24 268	29 987	37 426	46 981

平均指标时间数列中的各项指标数值也不能相加，因为加总后的结果不具有实际意义。

总量指标时间数列、相对指标时间数列和平均指标时间数列这3种时间数列，实际上并不是彼此孤立的。基于社会经济现象的复杂性和关联性，在很多情况下，为了准确、全面地分析事物发展变化过程，必须把这3种时间数列结合起来运用。

5.1.3 时间数列的编制原则

编制时间数列的主要目的，是通过对数列中一系列指标数值的研究分析，说明社会经济现象的发展变化过程及其规律性。因此，保证数列中各项指标数值的可比性，是编制时间数列的基本要求。编制时间数列应遵循以下基本原则。

1. 时期长短应相等

时间数列中各项指标数值的大小与其所属的时期长短有直接关系，因此，在同一个时期数列中，各指标数值所属时期长短要力求相等，否则时期长短不一，很难直接作出判断和比较。当然，对时间可比性的理解不能绝对化，有时为了特殊的研究目的或者信息资料的局限，也可以将时期不等的指标数值编制成时间数列，例如，表5-5中我国不同时期钢产量的动态数列。

表5-5 我国几个重要时期钢产量统计

年 份	1900—1949	1953—1957	1981—1985	1986—1990	1991—1995
钢产量/万吨	776	1 667	20 304	27 372	42 478

从表5-5中的数列可以看出，我国钢铁工业在新中国成立以后飞速发展，特别是改革开放以来取得了伟大成就。

对于时点数列，虽然指标数值的大小与时间间隔长短无直接关系，但也要尽量保持时点间隔的一致，以利于对比分析问题。

2. 总体范围要一致

总体范围是指被研究的社会经济现象所包括的地区范围、隶属范围、分组范围等。时间数列中各个指标所包括的总体范围应前后一致，若某一地区的行政区划发生变化，则该地区的总人口、土地面积、工业总产值、粮食总产量等指标的总体范围与以前有所不同，在编制有关指标的时间数列时，必须先将资料调整成总体范围前后一致，再进行动态分



析，这样才能正确说明所研究的问题。

3. 指标的经济内容要统一

统计指标数值反映一定的客观经济内容，因此，时间数列中的各项指标数值所包含的经济现象必须是同质的。例如，要反映某一国有工业企业发展变化的情况，可编制工业增加值数列，但不能将工业总产值与工业增加值混在一起编制。因为二者经济内容不同，没有可比性，否则会得出错误的结论。随着我国经济体制改革的不断深化，某些指标的经济含义也在变化，因此，保证各期指标经济内容的一致性就十分必要。

4. 各项指标数值的计算方法、计算口径和计量单位要一致

在指标名称及其经济内容一致的前提下，各项指标的计算方法、计算口径和计量单位也应当统一。例如，在编制不同时期第一、第二、第三产业增加值的时间数列时，应采用统一的不变价格计算，消除价格因素的影响，以便真实地反映各产业的发展情况；在研究企业劳动生产率的变动时，分子产量指标是用实物量还是用价值量，分母人数指标是用全部职工人数还是用生产工人人数，前后均应统一起来。

5.2 时间数列水平分析指标

5.2.1 发展水平

发展水平(time-series data)是时间数列中具体时间条件下的指标数值，具体反映某种社会经济现象在各个时期所达到的规模和发展的速度，也叫时间数列水平。它是计算其他动态分析指标的基础，一般用 a_i 表示。

发展水平可以表现为总量指标，也可以表现为相对指标或平均指标。

发展水平按在时间数列中的位置不同划分，有最初水平、最末水平和中间水平。在一个时间数列中，数列的第一项指标，称为最初水平，用符号 a_0 表示；最后一项指标称为最末水平，用符号 a_n 表示；其余就是中间发展水平，用符号 a_1，a_2，…，a_{n-1} 表示。

发展水平按在时间数列分析中的作用不同划分，有报告期水平与基期水平。将被研究时期的发展水平称为报告期水平或计算期水平，通常用 a_i 表示；将作为比较时期的发展水平称为基期水平或基础水平，用 a_0 或 a_{i-1} 表示。

在实际工作中，发展水平常用文字“增加到”、“增加为”、“降低到”、“降低为”表示。

5.2.2 平均发展水平

平均发展水平又称为序时平均数或动态平均数，它是将整个时间数列作为一个整体，从而反映这个整体的一般水平，即将动态数列不同时间上的发展水平加以平均而得到的平均数。

1. 序时平均数与一般平均数的区别与联系

序时平均数与前面所讲的一般平均数有共同之处，都是将现象总体的个体数量差异抽象化，反映社会经济现象的一般水平。但二者却有区别，主要表现在以下几个方面。

(1) 序时平均数是事物在不同时间上的数量差异的平均；算术平均数是总体各单位某一数量标志在同一时间上的数量差异的平均。

(2) 序时平均数是从动态上说明某一事物在不同时间上发展的一般水平；算术平均数是从静态上说明同一事物总体的不同单位在同一时间上的一般水平。

(3) 序时平均数是根据时间数列计算的；算术平均数是根据变量数列计算的。

在动态分析中，利用序时平均数(平均发展水平)分析社会经济现象的动态变化有很重要的作用：首先，它可以反映社会经济现象在一段时间内发展变化的一般水平，并对其作出概括的说明；其次，利用序时平均数可以消除社会经济现象在短期内所受波动的影响，便于在更广泛的范围内进行对比，可以观察现象的发展趋势；最后，利用序时平均数还可以对某一段时间内某一事物发展达到的一般水平进行不同单位、不同地区之间的比较。

2. 平均发展水平的计算方法

1) 根据总量指标时间数列计算平均发展水平

总量指标时间数列分为时期数列与时点数列两种。计算平均发展水平，需要根据时间数列不同特点采用不同的方法，现分别讨论如下。

(1) 根据时期数列计算平均发展水平。由于时期数列中的各项指标数值都是反映社会经济现象在一定时期内的过程总量，具有可加性，因此可以采用简单算术平均的方法计算平均发展水平，即将时期数列中研究范围内的各项指标数值之和除以时期项数得到。计算公式为

$$\bar{a}=\frac{a_1+a_2+\cdots+a_n}{n}=\frac{\sum a}{n} \tag{5-1}$$

式中：$\bar{a}$ 为发展水平；$a_i(i=1，2，3，\cdots，n)$为发展水平；n 为项数。

【实例 5.1】 根据表 5-1，现要求计算李先生的超市 2006—2012 年年平均销售额。

解：

在该例中，由于销售额是时期指标，所以可采用简单算术平均方法，即该超市 2006—2012 年年平均销售额为

$$\bar{a}=\frac{a_1+a_2+\cdots+a_n}{n}=\frac{\sum a}{n}$$

$$\approx 24.47(\text{百万元})$$

(2) 根据时点数列计算平均发展水平。由于时点数列中的各项指标数值都是社会经济现象在某一具体时点条件下的瞬间水平，要计算其平均数，就必须知道在每一时点上的指标数值，而事实上这是不可能的，所以在实际中都有一定的时间间隔。根据资料的掌握情况及排列情况，时点数列又可分为连续时点数列与间断时点数列。所谓连续时点数列，是在实践中，通常用“天”作为最小的时点单位，如果资料逐日登记，就将它看成连续时点



数列；如果资料是通过间隔一定时期登记取得的，通常是期初或期末(如月初或月末、季初或季末、年初或年末)登记取得的，就将它看成间断时点数列。

① 根据连续时点数列计算平均发展水平。对于连续时点数列分为两种情况。一是资料未经整理，逐日登记，又是逐日排列的，称为间隔相等的连续时点数列。

间隔相等的连续时点数列的平均发展水平的计算公式为

$$\bar{a}=\frac{\sum a}{n} \tag{5-2}$$

式中：a 为时点指标数值；n 为天数。

【实例 5.2】 某公司业务一科 9 月上旬每日出勤人数分别为 14、15、14、13、15、14、14、13、15、13。计算 9 月上旬平均每日出勤人数。

解：

$$\bar{a}=\frac{\sum a}{n}=\frac{14+15+14+13+15+14+14+13+15+13}{10}=14(\text{人})$$

二是数列中的各项指标并非逐日变动，在发生变动时才加以登记，称为间隔不等的连续时点数列。此时，资料经过整理，形成频数分布，即存在若干天指标数值相同，则需要以指标数值为变量，以持续天数为权数进行加权算术平均。

间隔不等的连续时点数列的平均发展水平的计算公式为

$$\bar{a}=\frac{\sum af}{\sum f} \tag{5-3}$$

式中：f 为各项指标的时间间隔，其余符号与前面公式相同。

【实例 5.3】 某企业 2013 年 3 月成品库存资料记录见表 5-6。

表 5-6 某企业 2013 年 3 月成品库存量

时　　间	1 日	10 日	15 日	25 日	26 日	28 日
库存量/台	100	240	400	300	200	200

现要求计算该企业 3 月成品的平均库存量。

解：

从表 5-6 中可以看出，库存量虽然表现为间隔资料(即不是逐日排列)，但实质上具有逐日的详细资料。例如，1 日库存量为 100 台，2～9 日虽没有列出，实质上 2～9 日的库存量没有变化，仍保持 100 台，直到 10 日入库 140 台，变为 240 台。在计算该企业 3 月平均成品库存量时，应以库存量为变量，以持续天数为权数进行加权算术平均，即

$$\bar{a}=\frac{\sum af}{\sum f}=\frac{100\times9+240\times5+400\times10+300\times1+200\times2+200\times4}{31}$$

$$\approx245.16(\text{台})$$

② 根据间断时点数列计算平均发展水平。对于间断时点数列分为两种情况。一是时点数列间隔相等，称为间隔相等的间断时点数列。

间隔相等的间断时点数列的平均发展水平的计算公式为

$$\bar{a}=\frac{\frac{a_1}{2}+a_2+a_3+\cdots a_{n-1}+\frac{a_n}{2}}{n-1} \tag{5-4}$$

式中：$n-1$ 为间断数目。

利用这种方法计算平均发展水平有一个前提条件，即假定现象在相邻两个时点之间的发展变动是均匀的。首先以每一小段的中间值代表该小段的平均水平，然后将各小段的平均水平再用简单算术平均法计算，得到整个被研究时期的平均发展水平。该方法称为“首尾折半法”或“首末折半法”。

【实例 5.4】 某公司 2013 年第一季度员工人数见表 5－7。

表 5－7　某公司 2013 年第一季度员工人数

时　　间	1 月 1 日	2 月 1 日	3 月 1 日	4 月 1 日
员工人数/人	150	160	156	162

计算第一季度平均员工人数。

解：

第一季度包括 1、2、3 共 3 个月，要求该季平均人数，则应先按月求出这 3 个月的平均人数，然后对各月平均人数进行平均即为所求。要计算月平均人数，严格地说，必须具有月内每日的人数资料。由前面讨论得知，在实践中，一般是通过期初、期末登记取得资料，而并非逐日登记。所以，只能采用假定的方法推算月平均人数，把月底的人数看成下月初的人数，并假定月中的变动是均匀的，于是便得到

$$月平均人数=\frac{月初人数+月末人数}{2}$$

$$1\ 月人数=\frac{150+160}{2}=155(人)$$

$$2\ 月人数=\frac{160+156}{2}=158(人)$$

$$3\ 月人数=\frac{156+162}{2}=159(人)$$

该公司第一季度平均员工人数为

$$\begin{aligned}\bar{a}&=\frac{(150+160)/2+(160+156)/2+(156+162)/2}{4-1}\\&=\frac{150/2+160+156+162/2}{3}\\&\approx 157.3(人)\end{aligned}$$

二是时点数列间隔不等，称为间隔不等的间断时点数列。

间隔不等的间断时点数列的平均发展水平的计算公式为

$$\bar{a}=\frac{\frac{a_1+a_2}{2}f_1+\frac{a_2+a_3}{2}f_2+\cdots+\frac{a_{n-1}+a_n}{2}f_{n-1}}{f_1+f_2+\cdots+f_{n-1}} \tag{5-5}$$



式中：f 为各时点的间隔长度。

利用这种方法计算平均发展水平，是以各期期末资料不全为条件，以资料欠缺的时点间隔长度为权数而采用的加权算术平均法。即先求各间隔内的平均数，然后求总平均数。

注意：上式中指标数值项数为 n，间隔为 $n-1$ 个；间隔之和 $f_1+f_2+\cdots+f_{n-1}$ 不能写成 $\sum f$。

【实例 5.5】 某公司 2012 年度职工人数见表 5－8。

表 5－8　某公司 2012 年度职工人数

时　　间	1月1日	5月1日	8月1日	11月31日	12月31日
职工人数/人	400	420	426	432	440

计算 2012 年度该公司年平均职工人数。

解：

$$\bar{a}=\frac{\frac{400+420}{2}\times 4+\frac{420+426}{2}\times 3+\frac{426+432}{2}\times 4+\frac{432+440}{2}\times 1}{4+3+4+1}$$

$$=\frac{5\ 061}{12}=421.75(\text{人})$$

由于根据间断时点数列计算平均发展水平是以一定程度假设性为前提的，即假设社会经济现象在各时点间是均匀变化的。因此，在实际中，间断时点数列的时间间隔不宜过长，否则准确度较差。

2)根据相对指标时间数列计算平均发展水平

相对指标时间数列是由具有相互联系的两个总量指标时间数列对比构成的，且作为对比基础的总量指标一般不相等。因此，根据相对指标时间数列计算平均发展水平时，其基本方法步骤是：首先，计算出分子数列和分母数列的平均发展水平；其次，将这两个平均发展水平进行对比。其基本计算公式为

$$\bar{c}=\frac{\bar{a}}{\bar{b}} \tag{5-6}$$

式中：$\bar{c}$ 为相对指标时间数列的平均发展水平；$\bar{a}$ 为分子数列的平均发展水平；$\bar{b}$ 为分母数列的平均发展水平。

在实际工作中，相对指标时间数列可以由两个时期数列对比形成，也可以由两个时点数列对比形成，还可以由一个时期数列与一个时点数列对比形成。各种数列性质不同，其计算相对指标时间数列的平均发展水平的方法也不一样，具体有 3 种情况。

第一种情况：根据两个时期数列对比所形成的相对指标时间数列计算平均发展水平。其计算公式为

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{\frac{\sum a}{n}}{\frac{\sum b}{n}}=\frac{\sum a}{\sum b} \tag{5-7}$$

【实例 5.6】 某商业企业 2013 年第二季度月销售额计划完成情况见表 5－9。

表 5－9 某商业企业第二季度销售额计划完成情况

时　间	4 月	5 月	6 月
实际完成销售额(a)/万元	125.6	136.7	197.8
计划完成销售额(b)/万元	115.0	128.0	176.0
销售额计划完成程度(c)	109.2%	106.8%	112.4%

计算第二季度销售额计划完成程度。

解：

销售额计划完成程度的时间数列是相对指标时间数列，是由实际销售额和计划销售额这两个时期数列进行对比计算出来的。由于时期数列的各项指标可以相加，反映更长时间内的现象总量，因此该商业企业第二季度平均销售额计划完成程度为

$$\bar{c}=\frac{\bar{a}}{\bar{b}}\times 100\%=\frac{\frac{\sum a}{n}}{\frac{\sum b}{n}}\times 100\%=\frac{\sum a}{\sum b}\times 100\%=\frac{125.6+136.7+197.8}{115.0+128.0+176.0}\times 100\%$$

$$=\frac{460.1}{419.0}\times 100\%\approx 109.8\%$$

第二种情况：根据两个时点数列进行对比所形成的相对指标时间数列计算平均发展水平。这种平均数一般是根据间隔相等的间断时点数列所组成的时间数列来计算的。其平均发展水平的计算公式为

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{\frac{a_1}{2}+a_2+a_3+\cdots+a_{n-1}+\frac{a_n}{2}}{\frac{b_1}{2}+b_2+b_3+\cdots+b_{n-1}+\frac{b_n}{2}} \tag{5-8}$$

【实例 5.7】 某投资咨询公司的职工人数见表 5－10。

表 5－10 某投资咨询公司 2012 年第三季度职工人数

月　份	6 月末	7 月末	8 月末	9 月末
投资咨询业务人数/人	645	670	695	710
全部职工人数/人	805	826	830	854
投资咨询业务人员占全部职工人数比重	80.1	81.1%	83.7%	83.1%

试计算第三季度投资咨询业务人员占全部职工人数的比重。

解：

上述资料是时间间隔相等的间断时点数列，要计算该公司第三季度投资咨询业务人员占全部职工人数的平均比重，必须先计算出第三季度平均投资咨询业务人员人数和平均全部职工人数，然后把它们相对比进行求解。



$$\bar{c}=\frac{\bar{a}}{\bar{b}}\times 100\%=\frac{\frac{a_1}{2}+a_2+a_3+\cdots+a_{n-1}+\frac{a_n}{2}}{\frac{b_1}{2}+b_2+b_3+\cdots+b_{n-1}+\frac{b_n}{2}}\times 100\%$$

$$=\frac{\frac{645}{2}+670+695+\frac{710}{2}}{\frac{805}{2}+826+830+\frac{854}{2}}\times 100\%=\frac{2\ 042.5}{2\ 485.5}\times 100\%\approx 82.18\%$$

第三种情况：根据一个时期数列与另一个时点数列进行对比所形成的相对指标时间数列计算平均发展水平。其计算公式为

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{(a_1+a_2+a_3+\cdots+a_{n-1}+a_n)/n}{\left(\frac{b_0}{2}+b_1+b_2+b_3+\cdots+b_{n-1}+\frac{b_n}{2}\right)/n} \tag{5-9}$$

【实例 5.8】 某超市第四季度商品销售额与月末库存额见表 5－11。

表 5－11 某超市第四季度商品销售额与月末库存额

月　　份	9 月	10 月	11 月	12 月
商品销售额(a)/万元	75	80	150	240
月末库存额(b)/万元	35	45	55	65
商品流转次数(c)/次	—	2	3	4

试计算第四季度平均每月商品流转次数。

解：

商品流转次数时间数列是一个相对指标时间数列，构成相对数列的分子是商品销售额时期数列，分母是月末库存额时点数列，所以应该分别按时期数列计算分子平均发展水平，按间隔相等的间断时点数列计算分母平均发展水平，然后将分子与分母对比，求得相对指标时间数列的平均发展水平。

第四季度平均每月商品流转次数为

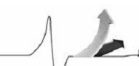
$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{(a_1+a_2+a_3+\cdots+a_{n-1}+a_n)/n}{\left(\frac{b_0}{2}+b_1+b_2+b_3+\cdots+b_{n-1}+\frac{b_n}{2}\right)/n}$$

$$=\frac{\frac{80+150+240}{3}}{\frac{\frac{35}{2}+45+55+\frac{65}{2}}{3}}=\frac{156.7}{50}\approx 3.13(\text{次})$$

3）根据平均指标时间数列计算平均发展水平

平均指标时间数列由一般平均数和序时平均数组成。由于这两种平均数各有不同的特点，因此，由它们组成的时间数列计算平均发展水平的方法也不相同。

第一种情况：根据一般平均数组成的时间数列来计算平均发展水平。一般平均数时间数列，实质上也是由两个总量指标时间数列对比所形成的，因此，其平均发展水平的计算与相对指标时间数列计算平均发展水平的方法相同。



第二种情况：根据序时平均数组成的时间数列来计算平均发展水平。根据间隔相等的序时平均数所组成的平均指标时间数列计算平均发展水平，可直接采用简单算术平均法计算，见式(5-10)。而由间隔不等的序时平均数所组成的平均指标时间数列计算平均发展水平，则以间隔长度为权数进行加权计算，见式(5-11)。

$$\bar{c}=\frac{\sum\bar{a}}{n} \tag{5-10}$$

$$\bar{c}=\frac{\sum\bar{a}f}{\sum f} \tag{5-11}$$

5.2.3 增长量

增长量又称增减量，是时间数列中报告期水平与基期水平之差，它说明社会经济现象在一定时期内增减变化的绝对量。其计算公式为

$$增长量=报告期水平-基期水平 \tag{5-12}$$

若计算结果为正值，表示增加量；若计算结果为负值，表示减少量。由于对比选择的基期不同，增长量可分为逐期增长量和累计增长量两种。

逐期增长量是报告期水平与上一期水平之差，表明各报告期水平较上一时期水平逐期增减变动的绝对数量，即“逐期增长量＝报告期水平－上一期水平”；累计增长量是报告期水平与某一固定基期水平之差，表明各报告期水平较某一固定时期水平增减变动的绝对数量，即累计增长量＝报告期水平－某一固定基期水平。以符号表示为

逐期增长量：a_1-a_0，a_2-a_1，a_3-a_2，…，a_n-a_{n-1}

累计增长量：a_1-a_0，a_2-a_0，a_3-a_0，…，a_n-a_0

逐期增长量与累计增长量之间有一定的数量关系，即累计增长量等于相应各个逐期增长量之和，用公式表示为

$$a_n-a_0=(a_1-a_0)+(a_2-a_1)+(a_3-a_2)+\cdots+(a_n-a_{n-1}) \tag{5-13}$$

【实例 5.9】 表 5－1 中李先生的超市 2006—2012 年年销售额的逐期增长量与累计增长量见表 5－12。

表 5－12 2006—2012 年年销售额的分析指标

年份		2006	2007	2008	2009	2010	2011	2012
销售额/百万元		5.6	11.4	17.1	24	31.3	38	43.9
增长量/百万元	累计	0	5.8	11.5	18.4	25.7	32.4	38.3
	逐期	—	5.8	5.7	6.9	7.3	6.7	5.9
发展速度	定基	100%	203.57%	305.36%	428.57%	558.93%	678.57%	783.93%
	环比	—	203.57%	150%	140.35%	130.42%	121.41%	115.53%
增长速度	定基	0	103.57%	205.36%	328.57%	458.93%	578.57%	683.93%
	环比	—	103.57%	50%	40.35%	30.42%	21.41%	15.53%
增长 1%的绝对值		—	0.056	0.114	0.171	0.24	0.313	0.38



此外，对于受季节因素影响较明显的社会经济指标，为了排除季节变化的影响，还可计算年距增长量。它是报告期为某月(季)水平与上年同月(季)水平之差，说明经济现象经过一年增减变化的绝对量。其计算公式为

$$年距增长量=本期发展水平-上年同期发展水平 \tag{5-14}$$

例如，该超市 2012 年年距增长量=43.9−38=5.9(百万元)。

5.2.4　平均增长量

平均增长量是指逐期增长量的简单算术平均数，它说明经济现象在一段较长时间内，每期平均增减变化的数量，其计算公式为

$$平均增长量=\frac{逐期增长量之和}{逐期增长量项数}=\frac{累计增长量}{数列项数-1} \tag{5-15}$$

【实例 5.10】 根据表 5-1 计算超市销售额在 2006—2012 年平均每年增长量。

解：

$$平均增长量=(5.8+5.7+6.9+7.3+6.7+5.9)/6\approx 6.38(百万元)$$

或

$$平均增长量=38.3/6\approx 6.38(百万元)$$

5.3　时间数列速度分析指标

5.3.1　发展速度

发展速度(rate of expansion)是反映客观现象发展变化快慢程度的相对指标，它是根据报告期水平与基期水平之比求得的，表明报告期水平已发展到基期水平的若干倍或百分之几。其计算公式为

$$发展速度=\frac{报告期水平}{基期水平} \tag{5-16}$$

由于对比所采用的基期不同，发展速度指标可分为定基发展速度和环比发展速度两种。定基发展速度是报告期水平与某一固定时期的水平之比，表明报告期水平对某一固定时期水平已发展到若干倍或百分之几，表明客观现象在较长时期内总的发展速度。其计算公式为

$$定基发展速度=\frac{报告期水平}{固定基期水平}\times 100\% \tag{5-17}$$

符号表示：$\frac{a_1}{a_0},\frac{a_2}{a_0},\frac{a_3}{a_0},\cdots,\frac{a_{n-1}}{a_0},\frac{a_n}{a_0}$

例如，表 5-12 中超市销售额的定基发展速度就是 2006—2012 年各期的发展水平分别与 2006 年的发展水平对比而得出的。

环比发展速度是报告期水平与前一期水平之比，说明报告期水平对前一期水平已发展到若干倍或百分之几，表明客观现象在两个相邻时期或时点上的发展速度。其计算公式为

$$环比发展速度=\frac{报告期水平}{前一期水平}\times 100\% \tag{5-18}$$

符号表示：$\frac{a_1}{a_0},\frac{a_2}{a_1},\frac{a_3}{a_2},\frac{a_4}{a_3},\cdots,\frac{a_n}{a_{n-1}}$

例如，表 5－12 中超市销售额的环比发展速度就是 2006—2012 年各期的发展水平分别与前一年的发展水平对比而得出的。

定基发展速度与环比发展速度虽然在说明客观现象时的侧重点有所不同，但它们之间存在着以下两种换算关系。

（1）定基发展速度等于相应时期内各个环比发展速度的连乘积，即

$$\frac{a_n}{a_0}=\frac{a_1}{a_0}\times\frac{a_2}{a_1}\times\frac{a_3}{a_2}\times\frac{a_4}{a_3}\times\cdots\times\frac{a_n}{a_{n-1}} \tag{5-19}$$

【实例 5.11】 由表 5－12 得出

2012 年的定基发展速度＝(2.035 7×1.50×1.403 5×1.304 2×1.214 1×1.155 3)×100%
≈783.99%

（2）相邻的两个定基发展速度相除，得出相应的环比发展速度，即

$$\frac{a_i}{a_0}\bigg/\frac{a_{i-1}}{a_0}=\frac{a_i}{a_{i-1}} \tag{5-20}$$

【实例 5.12】 由表 5－12 得出

2012 年的环比发展速度＝783.99/678.57×100%≈115.54%。

在实际工作中，可以根据以上两种换算关系，对发展速度进行相互推算。

为了消除季节因素对社会经济现象发展变化的影响，在计算月份或季度发展速度时，可选用上年同期作为对比的基期，计算年距发展速度。此外，还可以选用历史最高水平的时间作为对比的基期，以反映在报告期已经发展到或不及历史最高水平的程度。

5.3.2 增长速度

增长速度(rate of growth)又称增减速度，是反映客观现象增长快慢程度的相对指标，它是增长量与基期发展水平之比，用以说明报告期水平比基期水平增长了若干倍或百分之几。其计算公式为

$$\begin{aligned}增长速度&=\frac{增长量}{基期水平}\times100\%=\frac{报告期水平-基期水平}{基期水平}\times100\%\\&=发展速度-1\end{aligned} \tag{5-21}$$

从上式可以看出，当发展速度大于 1 或 100%时，增长速度为正值，表示现象的增长程度；当发展速度小于 1 或 100%时，增长速度为负值，表示现象的降低程度。

增长速度由于所采用的基期不同，也可分为定基增长速度和环比增长速度。

定基增长速度是累计增长量与固定时期发展水平之比，它表明客观现象从某一固定基期至报告期这一时期内的增长速度。其计算公式为

$$\begin{aligned}定基增长速度&=\frac{a_i-a_0}{a_0}\times100\%=\frac{a_i}{a_0}\times100\%-1\\&=定基发展速度-1\quad(i=1,\ 2,\ 3,\ \cdots,\ n)\end{aligned} \tag{5-22}$$

例如，表 5－12 中超市销售额的定基增长速度就是由 2006—2012 年累计增长量除以固定基期(2006 年)水平而求得，也可以由定基发展速度减去 1 而求得。



环比增长速度是逐期增长量与前一期发展水平之比，它表明客观现象逐期增长的速度。其计算公式为

$$\text{环比增长速度}=\frac{a_i-a_{i-1}}{a_{i-1}}\times100\%=\frac{a_i}{a_{i-1}}\times100\%-1$$
$$=\text{环比发展速度}-1\quad(i=1,\ 2,\ 3,\ \cdots,\ n)\tag{5-23}$$

例如，表5-12中超市销售额的环比增长速度就是由2006—2012年各年的逐期增长量除以前一期水平而求得，也可以由环比发展速度减去1而求得。

值得注意的是，定基增长速度和环比增长速度之间没有量的乘除关系，所以不能直接进行换算。增长速度可以计算年距增长速度，也可以计算报告期水平与历史最高水平的差异程度。

为了反映增长速度的实际效果，有时需要计算每增长1%的绝对值指标，它表明速度每增长1%，发展水平指标在绝对数上增长了多少。其计算公式为

$$\text{增长}1\%\text{的绝对值}=\frac{\text{逐期增长量}}{\text{环比增长速度}\times100}$$
$$=\frac{\text{前一期水平}}{100}\tag{5-24}$$

例如，表5-12中超市销售额每年增长1%的绝对值是不同的，2007年仅为0.056百万元，2012年则为0.38百万元。可见，运用增长1%的绝对值指标，实际上遵循了相对指标要与总量指标相结合运用的原则，把增长速度和增长量结合起来观察、分析问题是非常必要的。

5.3.3 平均发展速度

平均发展速度是各个时期环比发展速度的序时平均数，说明客观现象在一个较长的时间内逐年平均发展变化的平均程度。平均发展速度也是一种动态平均数，但不能照搬前面计算序时平均数的方法来计算。在实际统计工作中，平均发展速度的计算方法有水平法和累积法两种。

1. 水平法

水平法又称几何平均法。平均发展速度是总速度的平均，且各环比发展速度的连乘积等于总速度，因此，计算现象平均发展速度时不能用算术平均法，而要用水平法计算。其计算公式为

$$\bar{x}=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}\times100\%=\sqrt[n]{\prod x}\times100\%\tag{5-25}$$

式中：$\bar{x}$为平均发展速度；x为各期环比发展速度；n为环比发展速度的项数；$\prod$为连乘符号。

由于动态数列中定基发展速度等于各环比发展速度的连乘积，所以，计算平均发展速度的公式又可以表示为

$$\bar{x}=\sqrt[n]{\frac{a_1}{a_0}\times\frac{a_2}{a_1}\times\cdots\times\frac{a_n}{a_{n-1}}}\times100\%=\sqrt[n]{\frac{a_n}{a_0}}\times100\%\tag{5-26}$$

一定时期的定基发展速度即为总速度。如果用 R 表示总速度，则平均发展速度的公式还可以表示为

$$\bar{x}=\sqrt[n]{R}\times 100\% \tag{5-27}$$

以上计算平均发展速度的 3 个公式，虽然形式不同，但其实质与计算结果完全相同。计算平均发展速度，究竟采用哪个公式，主要取决于所掌握的资料。如果掌握的资料是各年的环比发展速度，用式(5-25)；如果掌握的资料是最初水平和最末水平或各期发展水平，用式(5-26)；如果已知本期的定基发展总速度，则用式(5-27)。

【实例 5.13】 已知超市 2006—2012 年各年的销售额环比发展速度分别为 203.57%、150%、140.35%、130.42%、121.41%和 115.53%，求年平均发展速度。

解：

$$\bar{x}=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}\times 100\%=\sqrt[n]{\prod x}\times 100\%\approx 140.94\%$$

在实例 5.13 中，已知超市销售额 2006 年为 5.6 百万元，2012 年为 43.9 百万元，其年平均发展速度为

$$\bar{x}=\sqrt[n]{\frac{a_n}{a_0}}\times 100\%\approx 140.94\%$$

在实例 5.13 中，已知超市 2006—2012 年的销售额总发展速度为 783.93%，其年平均发展速度为

$$\bar{x}=\sqrt[n]{R}\times 100\%\approx 140.94\%$$

计算结果表明，应用以上 3 种公式对同一现象计算平均发展速度，所得出的结果是一致的。(有时出现小数不一致的现象，是计算过程中四舍五入造成的。)

【实例 5.14】 如果以 2012 年超市销售额为基数，其后每年若能以 140.94%的速度发展，到 2015 年该超市销售额将达到多少？

解：

$$a_n=a_o\cdot\bar{x}^n=43.9\times 1.409\,4^3\approx 122.90(\text{百万元})$$

即按 140.94%的速度发展，到 2015 年该超市销售额将达到 122.90 百万元。

【实例 5.15】 假设每年以 140.94%的速度发展，到哪一年该超市销售额将达到 80 百万元？

解：

根据公式(5-27)，可推导出 $n=\dfrac{\ln R}{\ln\bar{x}}\approx 7.75(\text{年})$

即 2006 年后 7.75 年，亦即到 2013 年第三季度左右，该超市销售额将达到 80 百万元。

用水平法计算平均发展速度具有两个特点：第一，这种方法侧重于考察最末一期的发展水平；第二，这种方法不能准确反映中间水平的起伏状况。

从理论上讲，用水平法计算的平均发展速度，是对一定发展阶段各期环比发展速度的平均，受各个时期发展水平的影响；但从计算公式观察，它只突出了最初水平和最末水平的影响，不能全面反映在整个发展阶段各期发展快慢的差别。

因此，在运用这一指标时，应注意最初水平与最末水平是否受特殊因素影响。同时，要联系各期环比发展速度加以分析，既要看水平，也要看速度，必要时用分段平均发展速度补充总平均发展速度，以便对现象的发展作出全面客观的评价。



2. 累积法

累积法又称方程式法。这种方法的特点是：从最初水平 a_0开始，每期以平均发展速度 $\bar{x}$ 发展，从理论上计算的各期水平之和应等于实际各期水平 $a_i(i=1,2,\cdots,n)$之和，即

$$\bar{x}+\bar{x}^2+\cdots+\bar{x}^n=\frac{\sum_{i=1}^{n}a_i}{a_0} \tag{5-28}$$

该高次方程的正根就是所要求的平均发展速度 $\bar{x}$ 。一般可以借助计算机求解，在实际工作中也可查“平均增长速度查对表”来求得。具体步骤如下。

（1）计算各期定基发展速度之和，即求 $\sum_{i=1}^{n}a_i/a_0$ 。

（2）用 $\sum_{i=1}^{n}a_i/a_0$ 再除以时期数 n，判断平均增长速度的正负性，如果除得的结果大于 1，表示各期的定基发展速度平均在 100%以上，则该数列属于递增型；反之，除得的结果小于 1，表示各期的定基发展速度平均在 100%以下，则该数列属于递减型。

（3）根据各期定基发展速度之和查“平均增长速度查对表”，即可得平均增长速度和平均发展速度。

【实例 5.16】 某乡镇企业 2007 年固定资产投资额为 122.3 万元，2008—2012 年累积投资达 701.5 万元，计算该企业 2007—2012 年固定资产投资的平均发展速度。

解：

已知 $a_0=122.3$ 万元，$\sum_{i=1}^{n}a_i=701.5$ 万元

则得 $\sum_{i=1}^{n}a_i/a_0=701.5/122.3\approx5.7359$ 即 573.59%

且 $5.7359/5\approx1.14718>1$

所以该数列为递增型。

查表 5-13，在 $n=5$ 一栏内，找接近 573.59%的数字为 573.38%，该数所对应的平均每年增长速度为 4.6%，即该企业 2007—2012 年的固定资产投资平均增长速度为 46%，平均发展速度为 104.6%。

表 5-13 累积法查对

平均每年增长	各年发展水平总和为基期的百分数				
	1 年	2 年	3 年	4 年	5 年
⋮	⋮	⋮	⋮	⋮	⋮
4.4%	104.40	213.39	327.18	445.98	570.01
4.5%	104.50	213.70	327.82	447.07	571.69
4.6%	104.60	214.01	328.46	448.17	573.38
4.7%	104.70	214.32	329.09	449.26	575.08
⋮	⋮	⋮	⋮	⋮	⋮

用于计算平均发展速度的水平法和累积法各自具有不同的特点和侧重面，计算平均发展速度究竟采用哪种方法，应根据所研究现象的具体情况确定。一般情况下，如果所研究现象侧重于考察最末一期所达到的水平，则采用水平法计算平均发展速度，如产量、工资等；如果所研究现象侧重于考察一段较长时间内的总量，则宜采用累积法计算平均发展速度，如基建投资、占用土地等。

5.3.4 平均增长速度

平均增长速度说明客观现象在一个较长的时间内逐年平均增长变化的程度。平均增长速度和平均发展速度的关系为

$$平均增长速度=平均发展速度-1 \tag{5-29}$$

但在实际工作中，平均增长速度不能根据各期的增长速度或总增长速度来计算，而是通过先计算平均发展速度，再依据两者之间关系进行换算。

【实例 5.17】 某生产公司产品单位成本在 5 年环比增减速度见表 5－14，求平均增长速度。

表 5－14 某生产公司 2008—2012 年产品单位成本环比增减速度

年　份	2007	2008	2009	2010	2011	2012
环比增长速度	—	－1.0%	－1.5%	－2.0%	－1.8%	－0.5%

解：

首先利用发展速度与增长速度之间的关系，求出各年环比发展速度，然后据此计算平均发展速度，最后得出平均增长速度。

$$\bar{x}=\sqrt[5]{(-1.0\%+1)\times(-1.5\%+1)\times(-2.0\%+1)\times(-1.8\%+1)\times(-0.5\%+1)}\times100\%$$
$$\approx98.64\%$$

$$平均增长速度=98.64\%-1=-1.36\%$$

5.4 时间数列趋势分析预测

社会经济现象的发展变化是由许多错综复杂的因素共同作用的结果，在这些因素中，有的是系统因素，它对现象的发展起着决定性作用；有的是偶然因素，它对现象的发展只起到局部的、临时的、非决定性的作用。由于各种因素的作用大小和方向不同，使时间数列呈现出不同的变动形态。时间数列分析的任务就是要正确地确定时间数列的性质，对各构成因素进行分解和测定，并对未来的状态作出推断和预测。作为基本分析，一般来说，把时间数列的构成因素按性质和作用分为 4 类，即长期趋势、季节变动、循环变动和不规则变动。

(1) 长期趋势(trend)是指序列变动的总的方向性趋势。形成一个序列的长期趋势是带根本性因素的变动趋势。例如，人口增长、科学技术的不断进步，必然导致社会生产总量呈现增长变动的趋势。这种长期趋势，通常可以认为是由各种固定的因素作用于同一方向而形成的，这些固定的因素，既有随时间推移按直线变化的，也有呈曲线变动的。所



以，长期趋势是对未来状况进行预测和推断的主要依据。

（2）季节变动(seasonal effects)是指在一年以内，时间数列受自然季节和社会习俗等因素影响而发生的有规律的、周期性的变动，如服装的季节性供应，铁路、公路、航空等客运量在春节前后出现高峰，等等。引起季节变动的原因，既有自然因素，也有人为因素，如气候条件、节假日及风俗习惯等。季节变动的影响有以1年为周期的，也有以1日、1周、1月为周期的。认识和掌握季节变动，对于近期行动的决策具有重要作用。

（3）循环变动(cyclical effects)是指围绕着长期趋势出现的，具有一定循环起伏形态的变动。这种循环起伏与季节变动不同，循环的幅度和周期很不规则，周期长的可达几十年，周期短的一般也在3～5年。例如，社会经济的繁荣、危机、萧条、复苏的周期变动便是典型的中循环波动，其周期一般在8～9年。循环变动也不同于长期趋势，它所表现的不是单一方向的持续运动，而是涨落相间的波浪式发展。测定循环变动，掌握其发展变化规律，这对于人们认识事物，利用或克服其产生的影响具有重要的意义。例如，我国统计工作中开展的宏观经济预测、预警系统的研究，有助于及时发现经济波动的趋势。

（4）不规则变动(irregular fluctuations)是指客观现象由于突发事件或偶然因素引起的无规律性的变动，又称剩余变动或随机变动，如政治动荡、大的自然灾害、战争等。它们是时间数列中无法由上述3个因素解释的部分。不规则变动与时间无关。

把时间数列的这些构成因素与时间数列的关系用一定的数学关系式表示，就是时间数列的分解模式。分解时间数列的基本模式有加法模式和乘法模式。

设时间数列为y，长期趋势为T，季节变动为S，循环变动为C，不规则变动为I，则时间数列分解的加法模式为

$$y=T+S+C+I \tag{5-30}$$

时间数列分解的乘法模式为

$$y=T\times S\times C\times I \tag{5-31}$$

值得注意的是，时间数列分析并不能作为对前景预测的唯一依据。因为时间数列中的不规则变动本身就是难以预计的，特别是一些偶然性重大事故的出现，规律更难以预料，即便是较有规律的因素(T、S、C)，它们的变动规律也不一定就是过去历史的简单重复。因此，在利用时间数列分析对社会经济现象进行预测时，预测的时间不宜过长，并应注意对一些影响其发展的主要因素进行分析。

在实际工作中，需要根据研究对象的性质、研究的目的和所掌握的资料等具体情况确定采用哪一种模式进行分析。一般来说，在社会经济统计中，主要采用乘法模式。

5.4.1 长期趋势的测定与预测

长期趋势是研究某种现象在一个相当长的时期内持续向上或向下发展变动的趋势。例如，我国的国内生产总值在社会主义市场经济运作中，呈现不断上升的长期趋势。测定长期趋势的主要目的有：首先，把握现象的趋势变化；其次，从数量方面研究现象发展的规律性，探求合适趋势线，为进行统计预测提供必要条件；最后，测定长期趋势，可以消除原有动态数列中长期趋势的影响，以便更好地显示和测定季节变动。

反映现象发展的长期趋势有两种基本形式：一种是直线趋势，另一种是非直线趋势，

即曲线趋势。当所研究现象在一个相当长的时期内呈现比较一致的上升或下降的变动，如遵循一直线发展，则为直线趋势，可求出一条直线代表它，这条直线也可叫做趋势直线。趋势直线上升或下降，表示这种现象的数值逐年俱增或俱减，且每年所增加或减少的数量大致相同。所以直线趋势的变化率或趋势线的斜率基本上是不变的。而非直线趋势，其变化率或趋势线的斜率是变动的。

研究现象发展的长期趋势，就必须对原来的时间数列进行统计处理，一般称为时间数列修匀，即进行长期趋势测定。测定长期趋势常用的主要方法有时距扩大法、移动平均法(moving average)、最小平方法(least square method)。

1. 时距扩大法

时距扩大法是测定直线趋势的一种简单的方法。当原始时间数列中各指标数值上下波动，使现象变化规律表现不明显时，可通过扩大数列的时间间隔，对原资料加以整理，以反映现象发展的趋势。用时距扩大法修匀时间数列，既可用总量指标表示，也可用平均指标表示。前者仅适用于时期数列，后者既适用于时期数列，也适用于时点数列。

【实例5.18】 利用时距扩大法修匀指标。某公司2012年产量见表5-15。

表5-15 某公司2012年产量(月份)

月　份	1	2	3	4	5	6	7	8	9	10	11	12
产量/万件	50	45.8	52	51.3	50.9	55.7	54.6	58.5	57	59.9	61.2	62.1

从表5-15中可看出，该公司各月产量有上升的发展趋势，但月与月之间，存在着升降交替的现象。如果将各月的产量合并为按季计算的产量，即扩大时距，则新编的动态数列见表5-16。

表5-16 某公司2012年产量(季度)

季　度	第一季度	第二季度	第三季度	第四季度
产量/万件	147.8	157.9	170.1	183.2

在时距扩大修匀后的新数列中，产量的变动情况就呈现出明显的上升趋势。

应用时距扩大法修匀动态数列，应遵循下列基本要求。

(1) 同一数列前后时间间隔应当一致，以便于比较。

(2) 时间间隔的长短，应根据具体现象的性质和特点而定，如果时距扩大不够，就不能消除现象变动中的偶然因素；反之，如果时距过长，则修匀后整理出来的动态数列指标就少，会掩盖现象发展的具体趋势。

2. 移动平均法

移动平均法是对原有时间数列进行修匀，来测定其长期趋势的一种较为简单的方法。这个方法就是采用逐项递推移动的方法，分别计算一系列移动的序时平均数，形成一个新的派生的序时平均数动态数列，来代替原有的时间数列。在这个新的时间数列中，短期的



偶然因素引起的变动被削弱了，从而呈现出明显的长期趋势。

【实例 5.19】 表 5-17 为某企业 2003—2012 年的商品销售额。

表 5-17 某企业 2003—2012 年的商品销售额

年 度	商品销售额/亿元	3 项移动平均	5 项移动平均	4 项移动平均	4 项移动平均正位
2003	9.60	—	—	—	—
2004	10.66	11.26	—	12.14	—
2005	13.52	12.98	12.32	13.00	12.58
2006	14.76	13.78	13.20	13.84	13.42
2007	13.08	13.94	14.08	14.22	14.04
2008	14.00	14.04	15.04	15.10	14.66
2009	15.04	15.78	15.68	16.32	15.72
2010	18.28	17.1	16.8	17.5	16.92
2011	17.96	18.32	—	—	—
2012	18.70	—	—	—	—

（1）三项移动平均。

第一个平均数$=\frac{9.60+10.66+13.52}{3}=11.26$，对正第二项的原值。

第二个平均数$=\frac{10.66+13.52+14.76}{3}=12.98$，对正第三项的原值。

依次类推，边移动边平均，求得 3 项移动平均新数列共 8 项。

（2）5 项移动平均。

第一个平均数$=\frac{9.60+10.66+13.52+14.76+13.08}{5}\approx 12.32$，对正第三项的原值。

第二个平均数$=\frac{10.66+13.52+14.76+13.08+14.00}{5}\approx 13.20$，对正第四项的原值。

依次类推，边移动边平均，求得 5 项移动平均新数列共 6 项。

（3）4 项移动平均。

第一个平均数$=\frac{9.60+10.66+13.52+14.76}{4}\approx 12.14$，对正第二项和第三项的中间。

第二个平均数$=\frac{10.66+13.52+14.76+13.08}{4}\approx 13.00$，对正第三项和第四项的中间。

依次类推，边移动边平均，求得 4 项移动平均新数列共 7 项。

由于每个指标数列都和原动态数列错半期，无法直接进行对比，还必须进行一次正位平均。即再进行一次两项移动平均，这样新序时平均数数列的各期才能和原动态数列对准，形成新的 4 项正位平均指标数列共 6 项。

从表 5-17 的计算结果可以看出，移动平均可以使时间数列中短期的偶然因素弱化，整个数列被修匀得更加平滑，波动趋于平稳。

应用移动平均法分析长期趋势时，应注意下列问题。

第一，用移动平均法对原动态数列修匀，修匀程度的大小与原数列移动平均的项数多少有关。5 项移动平均比 3 项移动平均修匀程度更大些。也就是说，修匀的项数越多，效果越好，即趋势线越平滑。

第二，移动平均法所取的项数的多少，应视资料的特点而定。原有时间数列如有循环周期，则移动平均的项数以循环周期的长度为准。事实证明，当移动平均的时期长度等于周期长度或其整倍数时，它就能把周期的波动完全抹掉。当数列资料为季度资料时，可采用 4 项移动平均；若根据各年的月份资料，则应取 12 项移动平均，这样可消除所受季节变动的影响，能较为准确地揭示现象发展的长期趋势。

第三，移动平均法采用奇数项移动比较简单，一次即得趋势值，见表 5-17。采用偶数项移动平均时，由于偶数项移动平均数都是在两项中间位置，所以要将第一次移动的平均值再进行两项“移正平均”，得出移正值时间数列，以显示出现象变动趋势。由于偶数项移动平均比较复杂，因此，一般常以奇数为长度。

第四，移动平均后的数列，要比原数列项数减少。移动时采用的项数越多，虽能更好地进行修匀，但所得趋势值的项数就越少。一般情况下，移动平均项数与趋势值的项数关系为

$$趋势值项数=原数列项数-移动平均项数+1 \tag{5-32}$$

因此，为了便于看出现象的发展趋势，要视具体情况，以确定移动平均的项数。

3. 最小平方法

最小平方法研究现象的发展趋势，就是运用一定的数学模型，对原有的时间数列配合一条适当的趋势线进行修匀。根据最小平方法的原理，这条趋势线必须满足最基本的要求，即原有数列的实际数值与趋势线的估计数值的离差平方之和为最小。用公式表示为

$$\sum (y-y_c)^2 \to 最小值 \tag{5-33}$$

式中：y_c为趋势线的估计数值；y 为原有数列的实际数值。

长期趋势的类型很多，有直线型和曲线型，而最小平方法既可用于配合直线，也可用于配合曲线，所以它是分析长期趋势的十分普遍和理想的方法。下面主要介绍根据社会经济现象的基本趋势，如何用最小平方法配合直线方程、抛物线方程及指数曲线方程。

1）直线方程

如果现象发展的逐期增长量大体上相等，则可考虑配合直线趋势。直线方程的一般形式为

$$y_c=a+bt \tag{5-34}$$

式中：y_c为趋势线的估计数值；T 为代表动态数列的时间单位；a 和 b 为参数，分别代表直线的截距和斜率。

根据最小平方法的基本要求，即 $\sum (y-y_c)^2 \to$ 最小值和多元函数求极值的定理，可用求偏导数的方法，导出以下联立方程组：

$$\begin{cases} \sum y = na + b\sum t \\ \sum ty = a\sum t + b\sum t^2 \end{cases} \tag{5-35}$$



式中：n 为动态数列的项数，其他符号与前面公式相同。

解联立方程组可得 a 和 b 两个参数值：

$$b=\frac{n\sum ty-\sum t\sum y}{n\sum t^2-(\sum t)^2}$$

$$a=\frac{\sum y-b\sum t}{n}=\bar{y}-b\bar{t} \tag{5-36}$$

将求得的两个参数代入直线趋势模型中，便可得到与实际观察值相对应的趋势值。由此可以形成一条趋势直线，既可以认识现象的发展变化动态，还可以预测未来。

【实例 5.20】 某地区 2007—2012 年家用洗衣机产量见表 5－18。求出直线趋势方程，并预测 2013 年家用洗衣机产量。

表 5－18 某地区 2007—2012 年家用洗衣机产量

年　份	产量(y)/千台	逐期增长量/千台	年份序号(t)	t^2	ty	y_c/千台
2007	68	0	0	0	0	67.3
2008	71	3	1	1	71	71.38
2009	75	4	2	4	150	75.46
2010	79	4	3	9	237	79.54
2011	84	5	4	16	336	83.62
2012	88	4	5	25	440	87.7
合计	465	—	15	55	1 234	465

解：

根据表 5－18，可知

$$n=6,\ \sum t=15,\ \sum y=465,\ \sum t^2=55,\ \sum ty=1\,234$$

将这些数据资料代入上面联立方程，得

$$\begin{cases}465=6a+15b\\1\,234=15a+55b\end{cases}$$

解方程组得 $a=67.3$，$b=4.08$

配合出的直线趋势方程式为 $y_c=67.3+4.08t$

利用已拟合的直线趋势方程式，预测该地区 2013 年家用洗衣机可能达到的产量为

$$y_{2013}=67.3+4.08\times6=91.78(\text{千台})$$

为了计算方便，可巧妙地对时间 t 进行取值。当时间项数为奇数时，可假设 t 的中间项为 0，这时时间项依次排列为…，－3，－2，－1，0，1，2，3，…；当时间项数为偶数时，时间项依次排列为…，－5，－3，－1，1，3，5，…，这样，原点 O 实际上是在数列正中相邻两个时间的中点，保证了时间项间隔相等。以上两种设 t 的方法目的是使 $\sum t=0$，此时，上述联立方程组可简化为

$$\begin{cases}\sum y = na \\ \sum ty = b\sum t^2\end{cases} \tag{5-37}$$

即

$$\begin{cases}a = \dfrac{\sum y}{n} \\ b = \dfrac{\sum ty}{\sum t^2}\end{cases} \tag{5-38}$$

这样可以大大简化计算过程。因此，通常把这种计算称为简捷法运算。在用最小平方法拟合直线方程时，不论时间数列的时间项数是奇数还是偶数，都可以运用简捷法来计算。

【实例 5.21】 仍以表 5－18 中的资料为例，用简捷法进行计算，见表 5－19。

表 5－19 简捷法计算

年 份	产量 y/千台	逐期增长量/千台	年份序号 t	t^2	ty	y_c/千台
2007	68	0	－5	25	－340	67.3
2008	71	3	－3	9	－213	71.38
2009	75	4	－1	1	－75	75.46
2010	79	4	1	1	79	79.54
2011	84	5	3	9	252	83.62
2012	88	4	5	25	440	87.7
合计	465	—	15	70	143	465

解：

依据表 5－19 中的资料，直线趋势模型的 a、b 两参数值为

$$\begin{cases}a = \dfrac{\sum y}{n} = \dfrac{465}{6} = 77.5 \\ b = \dfrac{\sum ty}{\sum t^2} = \dfrac{143}{70} \approx 2.04\end{cases}$$

将 a、b 两值代入直线趋势模型，得

$$y_c = 77.5 + 2.04t$$

预测该地区 2013 年家用洗衣机可能达到的产量($t=7$)为

$$y_c = 77.5 + 2.04 \times 7 = 91.78(\text{千台})$$

可见，用简捷法计算的各年趋势值和用一般计算方法计算的各年趋势值结果相同，以此预测的 2013 年该地区家用洗衣机可能达到的产量的结果也相同。

2） 抛物线方程

如果现象发展的逐期增长量的增长量(各期的二级增长量)大体相同，则可考虑曲线趋



势——配合抛物线方程。抛物线的一般方程为

$$y_c=a+bt+ct^2 \tag{5-39}$$

上述抛物线方程式中，有 a、b、c 3 个未定参数，根据最小平方法的要求，同样用求偏导数的方法，导出以下由 3 个方程组成的联立方程组：

$$\begin{cases}\sum y = na + b\sum t + c\sum t^2 \\ \sum ty = a\sum t + b\sum t^2 + c\sum t^3 \\ \sum t^2 y = a\sum t^2 + b\sum t^3 + c\sum t^4\end{cases} \tag{5-40}$$

同样，为了计算方便，可以通过假设 t，使 $\sum t=0$，$\sum t^3=0$，则上列联立方程组可简化为

$$\begin{cases}\sum y = na + c\sum t^2 \\ \sum ty = b\sum t^2 \\ \sum t^2 y = a\sum t^2 + c\sum t^4\end{cases} \tag{5-41}$$

对上述标准联立方程组求解，就可得到 a、b、c 的值，再将这 3 个参数代入抛物线方程，即为所求的抛物线方程模型。它同样可以帮助认识事物发展变化的规律，预测事物的未来。

3）指数曲线方程

如果现象发展的环比发展速度或环比增长速度大体相同，则可考虑用曲线趋势——配合指数曲线方程。指数曲线的一般方程为

$$y_c=ab^t \tag{5-42}$$

式中：a 为动态数列的基期水平；b 为现象的一般发展速度；t 为动态数列的时间；a、b 为未定参数。

进行指数曲线拟合时，一般是将指数方程通过取对数转化成直线方程，然后按直线方程办法确定参数，再对直线方程求得的结果查“反对数表”还原。

先对上述方程式两边各取对数，得

$$\lg y_c=\lg a+t\lg b$$

设：$Y=\lg y_c$，$A=\lg a$，$B=\lg b$

则 $Y=A+Bt$

根据最小平方法原理，推导出标准方程组为

$$\begin{cases}\sum Y = nA + B\sum t \\ \sum tY = A\sum t + B\sum t^2\end{cases} \tag{5-43}$$

同样设 t，使 $\sum t=0$，则此联立方程组可简化为

$$\begin{cases}\sum Y = nA \\ \sum tY = B\sum t^2\end{cases} \tag{5-44}$$

对上述联立方程组求解，即可求出 A 和 B 的值。由于 A 和 B 为对数值，需要查“反

对数表”求得 a 和 b 的值。把参数代入指数曲线模型，可以求得动态数列中与各观察值(实际值)对应的趋势值，由此形成一条趋势线。

5.4.2 季节变动的测定与预测

在一个动态数列中，除存在长期趋势外，往往还存在季节变动。所谓季节变动，是指一年以内的、周期性的有规律的变动。它既包括狭义的季节变动(如一年中春夏秋冬季节更替)，也包括广义的季节变动(如以1个月、1周甚至1日为周期的规律性变化)。例如，夏天汗衫、背心、冷饮的销售量就高于其他季节；冬天围巾、皮衣、取暖器的销售量就比较大；铁路客运量以春节前后为高峰；超市在周末销售额高于工作日销售额，等等。在经济生产中，季节变动会引起设备和劳动力使用的不平衡、原料供应不足、运输量不够，这给生产和人们生活带来某些影响。季节变动的原因，既有自然因素又有社会因素。研究季节变动主要是为了认识它、掌握它，合理利用季节变动的规律，以便为高效率组织工作、安排人民经济生活提供科学依据。

测定季节变动的方法很多，从其是否考虑受长期趋势的影响看，有两种方法：一种是不考虑长期趋势的影响，直接根据原始的时间数列计算，常用的方法是按月平均法；另一种是根据剔除长期趋势影响后的数列资料来计算，常用的方法是移动平均趋势剔除法。不管使用哪种方法计算季节变动，都需要用3年或更多年份的分月或分季资料作为基本数据进行计算分析，这样才能较好地消除偶然因素的影响，使季节变动的规律性更切合实际。

1. 按月(季)平均法

按月平均法亦称按季平均法。若是月资料就是按月平均；若是季资料则是按季平均。其计算的一般步骤如下。

(1) 列表。将各年同月(季)的数值列在同一栏内。

(2) 将各年同月(季)数值加总，并求出月(季)平均数。

(3) 将所有月(季)数值加总，求出总的月(季)平均数。

(4) 求季节比率(或季节指数)，其计算公式为

$$\text{季节比率}=\frac{\text{各月平均数}}{\text{全期各月平均数}}\times 100\% \tag{5-45}$$

如果计算出来的某月(季)的季节比率大于100%，则表明现象在该月(季)为旺季；若某月(季)的季节比率小于100%，则为淡季；如果等于100%，则表明该月不受季节变动的影响。

【实例5.22】 某商业企业毛线销售情况见表5-20，计算季节比率。

解：

第一，计算月(季)平均数；如1月为(400+400+460)/3=420，其余月份类推。

第二，计算总的月(季)平均数；把3年36个月的资料全部相加后平均，或用12个月平均数求总平均数，即(222.66+227.84+255.00)/3=235.17。

第三，计算各月平均数与总平均数的对比值，得出各月季节比率。例如，1月为420÷235.17×100%≈178.6%，其余月份类推。

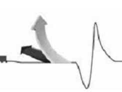


表5－20　某商业企业毛线销售季节变动

月　　份	第一年/kg	第二年/kg	第三年/kg	3年平均/kg	季节比率
1	400	400	460	420	178.6%
2	420	420	480	440	187.1%
3	300	260	340	300	127.6%
4	180	160	200	180	76.5%
5	140	160	180	160	68.0%
6	120	140	100	120	51.0%
7	100	98	120	106	45.1%
8	82	76	100	86	36.6%
9	160	180	200	180	76.5%
10	170	180	220	190	80.8%
11	220	240	200	220	93.6%
12	380	420	460	420	178.6%
合计	2 672	2 734	3 060	2 822	1 200.0%
平均	222.66	227.84	255.00	235.17	100.0%

计算结果表明，毛线销售呈明显的季节变动。每年12月至第二年3月为销售旺季，其中2月季节比率高达187.1%而进入顶峰；从4月开始进入淡季，一直延续到11月，其中8月季节比率仅有36.6%而降到低谷。据此，企业可以预测未来的季节变动，组织好毛线的经营。

按月平均预测法的计算公式为

各月(季)预测值＝上年各月(季)平均值×各月(季)的季节比率　　(5-46)

假定第四年该企业的毛线销售量计划为3 000kg，则

2月毛线销售的预测值：(3000÷12)×187.1%＝467.75(kg)

8月毛线销售的预测值：(3000÷12)×36.6%＝91.5(kg)

按月(季)平均法的优点是计算简便、容易理解，而缺点是不够精确，因为它没有消除时间数列中长期趋势的影响。要解决这一问题，就要用另一种方法，即趋势剔除法。

2. 移动平均趋势剔除法

移动平均趋势剔除法就是按移动平均法来剔除长期趋势的影响，再计算季节变动的方法。

【实例5.23】 某公司连续16个季度的商品销售量见表5－21，采用趋势剔除法计算其季节比率。

表 5－21　某公司商品销售量及计算

年　份	季度	销售量/万件	4 项移动平均	正位移动平均趋势值(y_c)	剔除趋势值(y/y_c)
2009	3	15	—	—	—
	4	18	12.0	—	—
2010	1	6	12.5	12.25	48.98
	2	9	13.0	12.75	70.59
	3	17	13.5	13.25	128.30
	4	20	14.0	13.75	145.45
2011	1	8	14.5	14.25	56.14
	2	11	15.0	14.75	74.58
	3	19	15.5	15.25	124.59
	4	22	16.0	15.75	135.21
2012	1	10	16.5	16.25	61.54
	2	13	17.0	16.75	77.61
	3	21	17.5	17.25	121.74
	4	24	18.0	17.75	135.21
2013	1	12		—	—
	2	13	—	—	—

解：

首先，采用 4 项(因为是季度资料)移动平均法，求出移动平均趋势值 y_c，见表 5－21。由于采用 4 项移动，其平均值对正两季中间，一般无法直接比较，所以应再进行一次两项移动平均，就得出对准各季的移正平均值，即表 5－21 中最后一栏的趋势值 y_c。

其次，从原数列剔除已测定的长期趋势变动。剔除趋势值的方法有两种：一是实际观测值 y 减去趋势值 y_c；另一种是实际观测值 y 除以趋势值 y_c。在此以第二种为例列表计算，见表 5－22。

原数列是由 16 个季度资料构成的，因为采用 4 项移动平均法，所以由趋势值形成的新数列只有全时期中间的 12 个季度的数字，即表 5－21 中 2010—2012 年的 12 个季度的数字。

再次，根据剔除趋势值所得的结果分别计算季节比率。计算季节比率的主要依据就是各年各季的 y/y_c 的百分比数字。将表 5－21 中各年间季的百分比数字重新排成一列，分别求出各季的平均数。计算结果见表 5－22。



表 5-22 趋势剔除法季节比率计算

年 份	第一季度	第二季度	第三季度	第四季度	合计
2010	48.98%	70.59%	128.30%	145.45%	393.32%
2011	56.14%	74.58%	124.59%	139.68%	394.99%
2012	61.54%	77.61%	121.74%	135.21%	396.10%
合计	166.66%	222.78%	374.63%	420.34%	1 184.41%
平均数	55.55%	74.26%	124.88%	140.11%	394.80%
季节比率	56.28%	75.24%	126.52%	141.96%	400.00%

最后，由于季节比率是围绕100%上下波动的，按4个季度计算的季节比率之和应为400%(按月的季节比率之和应为1 200%)；否则就应计算调整系数设法调整。调整系数的计算公式为

$$调整系数=400\%(1\ 200\%)/调整前各季(月)季节比率之和 \tag{5-47}$$

$$调整后的季节比率=调整前各季(月)季节比率\times调整系数 \tag{5-48}$$

在本例中，由于各季百分比数值的平均数之和为394.80%，因此需要校正。调整系数=400/394.80≈1.013 17。以这一系数分别乘各季的平均数，即得表5-22中的季节比率。例如，第一季度调整后的季节比率为55.55%×1.013 17≈56.28%；其余季度以此类推。

季节比率说明现象在剔除了长期趋势影响后，一年中各季的波动程度，由于是以百分比表示的典型波动，所以也可以用于季节比率的预测。

必须指出，某些现象的季节变动并非是永恒的规律，随着科技的进步和人们生活习惯的改变，某些社会现象的季节变动会被削弱甚至完全消失。例如，鸡在冬天通常不下蛋，但在现代化的养鸡场，用电灯代替阳光延长白昼，室内设有空调，饲料又有专门的配方，讲究营养，鸡在冬天照样下蛋。因此蛋产量的季节变动规律发生了改变。

在实际中，凡是在短期内现象有周期性的规律变动，都可拓广称为季节变动，因此季节变动的分析方法在日常经济生活中有宽广的应用空间。例如，一周内哪天公园的游客最多；超市的销售额在星期几最高；一天内哪些时间(小时)交通最拥挤；什么时段商场客流量最大；等等。

本章小结

时间数列是将某一统计指标的数值按时间的先后顺序排列起来所形成的数列，由现象所属时间和各个统计指标数值两个基本要素构成，可分为总量指标、相对指标和平均指标时间数列3种类型。可比性是编制时间数列的基本原则。

时间数列水平指标包括发展水平、平均发展水平、增长量和平均增长量4种。

平均发展水平与静态平均数相区别，但也有共同之处。在根据总量指标时间数列计算平均发展水平时，时期数列的计算较为简单，而时点数列的计算较为复杂。它分为4种计算形式，主要是间隔相等(不等)的连续时点数列和间隔相等(不等)的间断时点数列。间隔

相等的间断时点数列的平均发展水平最常用，间隔不等的间断时点数列的平均发展水平主要是在专门调查分析中使用；相对指标时间数列的平均发展水平的基本出发点是分子数列的平均水平与分母数列的平均水平对比。由静态平均数组成时间数列，其平均发展水平与相对指标时间数列平均发展水平的计算方法相同；由序时平均数组成的时间数列，其发展水平视具体情况分别采用简单平均法和加权算术平均法计算。

增长量分为累计增长量和逐期增长量两种。平均增长量是对逐期增长量求的简单算术平均数。

时间数列的速度指标包括发展速度、增长速度、平均发展速度和平均增长速度4种。

发展速度与增长速度均有定基和环比之分。增长速度由发展速度减1得到。为了把速度指标与水平指标结合起来说明问题，通常运用增长1%的绝对值。

平均发展速度有水平法和累积法两种计算方法。水平法侧重考察最末发展水平，累积法侧重考察全期总水平。实际工作中水平法应用广泛。平均增长速度由平均发展速度减1得到。

影响时间数列的因素有长期趋势、季节变动、循环变动和不规则变动4种。测定长期趋势的方法有时距扩大法、移动平均法和最小平方法。最小平方法是依据时间数列的观察值与趋势值的离差平方和为最小值的基本要求拟合的一种趋势模型。最小平方法可以采用简捷法计算，即取 $\sum t=0$。

季节变动的分析测定方法，按是否消除长期趋势分为按月(季)平均法和移动平均趋势剔除法。前者不考虑长期趋势的变动，后者则剔除长期趋势的影响。



国际地位稳步提高 国际影响持续扩大

——“十一五”经济社会发展成就系列报告之十七

“十一五”时期，我国经济社会发展又迈上了一个新台阶，国际地位稳步提高，国际影响持续扩大。5年间，我国经济总量从世界第5位跃居世界第2位；货物出口额从世界第3位跃居第1位；外商直接投资从世界第4位跃居第2位，其他许多经济指标在国际上的排位也有显著进步。在这不平凡的5年里，我国保持了经济平稳较快发展，并积极开展国际合作，对推动世界经济复苏作出了重要贡献，产生了广泛而积极的影响。

1. 经济保持较快增长，经济总量跃居世界第2位

“十一五”期间，中国经济年均增长11.2%，远高于同期3.5%的世界平均水平。相关内容见表5-23。

“十一五”期间，国内生产总值(GDP)居世界的位次从2005年的第5位提升到2006的第4位、2007年的第3位，2010年首次超过日本，成为世界第二大经济体。中国GDP占世界的比重逐年上升，从2005年的5%提高到2010年的9.5%。同时，中国与美国的差距逐步缩小，相当于美国GDP的比例从2005年的17.9%上升至2010年的40.2%。相关内容见表5-24。



表5-23 2006—2010年世界主要国家和地区经济增长率比较

国家和地区 \ 年份	2006	2007	2008	2009	2010	2006—2010年平均增长率
世界总计	5.2%	5.3%	2.8%	−0.6%	5.0%	3.5%
美国	2.7%	2.0%	0.0%	−2.6%	2.8%	0.9%
欧元区	3.0%	2.9%	0.5%	−4.1%	1.8%	0.8%
日本	2.0%	2.4%	−1.2%	−5.2%	4.3%	0.4%
中国	12.7%	14.2%	9.6%	9.2%	10.3%	11.2%
中国香港	7.0%	6.4%	2.2%	−2.8%	6.0%	3.7%
韩国	5.2%	5.1%	2.3%	0.2%	6.1%	3.7%
新加坡	8.6%	8.5%	1.8%	−1.3%	15.0%	6.4%
南非	5.6%	5.5%	3.7%	−1.8%	2.8%	3.1%
印度	9.7%	9.9%	6.4%	5.7%	9.7%	8.2%
俄罗斯	8.2%	8.5%	5.2%	−7.9%	3.7%	3.4%
巴西	4.0%	6.1%	5.1%	−0.2%	7.5%	4.5%

表5-24 2005—2010年中国国内生产总值(GDP)居世界位次变化

年份	位次	国内生产总值/亿美元	占世界比重	相当美国的比例
2005	5	22 569	5.0%	17.9%
2006	4	27 129	5.5%	20.3%
2007	3	34 942	6.3%	24.9%
2008	3	45 200	7.4%	31.5%
2009	3	49 847	8.6%	35.3%
2010	2	58 791	9.5%	40.2%

2. 人均国民总收入大幅提高

“十一五”期间，我国人均国民总收入(GNI)大幅提高。2009年我国人均GNI达到3 650美元，比2005年增加了1.1倍，居世界第125位。我国人均GNI相当于世界平均水平的比例从2005年的24.8%，提高到2009年的41.8%，5年内差距缩小了17个百分点。

3. 主要工农业产品产量稳居世界前列

“十一五”期间，我国主要工业产品产量稳居世界前列，其中粗钢、硬煤、水泥和化肥产量居世界第1位；发电量保持世界第2位，仅次于美国；原油产量居世界第4位。相关内容见表5-25。

表 5-25　2005—2009 年我国主要工业产品产量居世界位次

工业产品	2005 年		2006 年		2007 年		2008 年		2009 年	
	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次
粗钢	34 806	1	42 145	1	48 701	1	49 790	1	56 640	1
硬煤	181 069	1	205 546	1	230 675	1	258 566	1	298 399	1
原油	18 068	5	18 472	5	18 671	5	19 002	5	18 960	4
发电量	23 974	2	27 494	2	31 806	2	33 923	2	36 213	2
水泥	106 885	1	123 677	1	135 413	1	138 286	1	163 726	1
化肥	4 149	1	4 590	1	5 129	1	5124	1		

“十一五”期间，我国主要农业产品产量都继续增加，其中谷物、肉类、籽棉、花生、茶叶和水果产量均稳居世界第 1 位；大豆和甘蔗产量分别稳居世界第 4 位和第 3 位；油菜籽产量 2008 年居世界第 2 位，其他年份均居世界第 1 位。相关内容见表 5-26。

表 5-26　2005—2009 年我国主要农业产品产量居世界位次

农业产品	2005 年		2006 年		2007 年		2008 年		2009 年	
	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次	产量/万吨	位次
谷物	42 937	1	45 279	1	45 781	1	48 005	1	48 368	1
肉类	7 119	1	7 269	1	7 042	1	7 451	1	7 821	1
籽棉	1 714	1	2 024	1	2 287	1	2 250	1	2 300	1
大豆	1 635	4	1 550	4	1 273	4	1 555	4	1 450	4
花生	1 440	1	1 281	1	1 308	1	1 434	1	1 334	1
油菜籽	1 305	1	1 097	1	1 057	1	1 210	2	1 350	1
甘蔗	8 758	3	9 331	3	11 373	3	12 492	3	11 375	3
茶叶	95	1	105	1	118	1	128	1	132	1
水果①	9 040	1	9 677	1	10 243	1	10 962	1	11 414	1

注：①不包括瓜类。

4. 货物进出口总额跃居世界第 2 位，服务贸易总额跃居世界第 4 位

“十一五”期间，我国货物进出口占国际市场份额持续扩大。2010 年，我国货物进出口总额 29 728 亿美元，比 2005 年增加了 1.1 倍。2009 年，我国货物进出口总额超过了德国，跃居世界第 2 位，占世界的比重从 2005 年的 6.7%提高到 8.8%。“十一五”期间，我国服务贸易在世界的排位不断提高。2009 年，居世界的位次从 2005 年第 8 位上升到第 4 位。

5. 外商直接投资跃居世界第 2 位，对外直接投资跃居世界第 5 位

2010 年，中国外商直接投资(FDI)为 1 057 亿美元，居世界的位次从 2005 年的第 4 位提高到第 2 位，占世界的比重从 2005 年的 7.3%提高到 9.4%，中国已成为世界最具吸引力的外商直接投资国之一。



6. 国际旅游规模持续扩大

2010年，中国出境旅游人数为5 739万人次，比2005年增长85%，入境(过夜)旅游人数为5 566万人次，比2005年增长18.9%。据世界银行统计，2008年中国出境旅游人数居世界第5位，比2005年提高1位，占世界总量的比重从2005年的3.4%提高到4.5%；2008年入境(过夜)旅游人数居世界第4位，占世界总量的5.7%。

2010年，中国国际旅游收入为458亿美元，比2005年增长44%。据世界银行统计，2008年中国国际旅游收入占世界总收入的3.9%，居世界第7位；中国国际旅游支出占世界总支出的4%，居世界位次从2005年的第7位上升到第5位。

7. 人文发展指数持续提高，居民生活条件不断改善

据联合国开发计划署测算，2010年中国人文发展指数(HDI)为0.663，超过世界平均水平(0.624)和中等国家水平(0.592)，在169个国家中居89位，比2005年提高8位。

“十一五”期间，居民生活水平不断提高，恩格尔系数持续下降，居住条件得到明显改善。2008年我国享有清洁引用水源人口占总人口比重为89%，达到中等收入国家水平(88%)；中国享有卫生设施人口占总人口比重为55%，接近中等收入国家水平(57%)。

8. 国际竞争力显著增强，国际影响力不断扩大

“十一五”期间，中国在机构、基础设施、宏观经济环境、健康与教育培训、商品市场效率等方面的国际竞争力都有了显著增强。据世界经济论坛《2010—2011年全球竞争力报告》测算，中国国际竞争力在139个国家和地区中排名第27位，而2005年在被评估的117个国家和地区中仅排名第49位。

据美国《财富》杂志统计，中国(包括中国台湾)进入世界500强企业数量从2005年的23家增加到2009年的54家，中国(包括中国台湾)进入500强企业的利润总额占世界比重从2005年的3.7%提高到2009年13.5%。

(资料来源：http://www.chinanews.com/cj/2011/03-24/2929155.shtml.)



习题与实训

一、单项选择题

1. “某公司顾客满意保障体系带来的效果如下：一是客户满意度大大提高，由过去的80%提高到96%，进而提高到现在的99%；二是质量问题重复投诉率下降，由过去的29%下降到10%，进而下降到现在的8%。”在这段分析报告中使用的数字属于时间数列中的发展水平中的(　　)指标。

A. 最初水平　　B. 中间水平

C. 最末水平　　D. 最初水平、中间水平、最末水平

2. 设2008—2012年各年的环比增长速度为6%、7%、8%、9%和10%，则平均增长速度为(　　)。

A. $\sqrt[5]{6\%\times7\%\times8\%\times9\%\times10\%}\times100\%$

B. $\sqrt[5]{106\%\times107\%\times108\%\times109\%\times110\%}\times100\%$

C. $\sqrt[5]{106\%+107\%+108\%+109\%+110\%}\times100\%$

D. $\sqrt[5]{106\%\times107\%\times108\%\times109\%\times110\%}\times100\%-100\%$

3. 已知某企业1月、2月、3月、4月的平均职工人数分别为190人、195人、193人和201人。则该企业一季度的平均职工人数的计算方法为(　　)。

A. $\dfrac{190+195+193+201}{4}$　　B. $\dfrac{190+195+193}{3}$

C. $\dfrac{\frac{190}{2}+195+193+\frac{201}{2}}{4-1}$　　D. $\dfrac{\frac{190}{2}+195+193+\frac{201}{2}}{4}$

4. 下面属于间隔不等的时点数列求序时平均数的公式的是(　　)。

A. $\bar{a}=\dfrac{\left(\frac{a_1+a_2}{2}\right)f_1+\left(\frac{a_2+a_3}{2}\right)f_2+\cdots+\left(\frac{a_{n-1}+a_n}{2}\right)f_{n-1}}{\sum f}$

B. $\bar{a}=\dfrac{\sum a}{n}$

C. $\bar{a}=\dfrac{\sum af}{\sum f}$

D. $\bar{a}=\dfrac{\frac{a_1}{2}+a_2+\cdots+\frac{a_n}{2}}{n-1}$

5. 定基增长速度与环比增长速度的关系是(　　)。

A. 定基增长速度是环比增长速度的连乘积

B. 定基增长速度是环比增长速度之和

C. 定基增长速度是各环比增长速度加1后的连乘积减1

D. 定基增长速度是各环比增长速度减1后的连乘积减1

6. 以2000年为基期，2013年为报告期，计算经济现象的平均发展速度，应开(　　)。

A. 12次方　　B. 13次方　　C. 11次方　　D. 14次方

7. “高水平难以高速度，而低水平却可以高速度”，下面(　　)指标可以解释其关键内涵。

A. 平均发展速度　　B. 平均增长量

C. 平均发展水平　　D. 增长1%绝对值

8. 某地区2006—2012年粮食产量(万吨)分别是250、258、266、273、279、286、294，以2006年为时间t的原点(即$t=0$)，则不需计算可知下面趋势方程正确的是(　　)。

A. $y_c=67.3+14.95t$　　B. $y_c=67.3+7.18t$

C. $y_c=250.75-7.18t$　　D. $y_c=250.75+7.18t$

9. 已知某地区建筑业产值第一、第二、第三季度的季节比率为74.00%、104.50%、121.00%，则由此可推测出第四季度是(　　)。

A. 旺季　　B. 淡季　　C. 平季　　D. 难下结论

10. 已知某企业1月、2月、3月、4月的月初职工人数分别为190人、195人、193人和201人。则该企业一季度的平均职工人数的计算方法为(　　)。



A. $\frac{190+195+193+201}{4}$

B. $\frac{190+195+193}{3}$

C. $\frac{\frac{190}{2}+195+193+\frac{201}{2}}{4-1}$

D. $\frac{\frac{190}{2}+195+193+\frac{201}{2}}{4}$

二、多项选择题

1. 时间数列的编制原则有(　　)。

A. 时期长短应相等

B. 总体范围要一致

C. 指标的经济内容要统一

D. 各项指标数值的计算方法、计算口径和计量单位要一致

E. 时间数列的两个要素均要“可比”

2. 按统计指标表现的形式看，时间数列可分为(　　)。

A. 总量指标时间数列

B. 相对指标时间数列

C. 平均指标时间数列

D. 时期指标时间数列

E. 时点指标时间数列

3. 下面是时期数列的有(　　)。

A. 我国近几年的耕地总面积

B. 我国历年新增人口数

C. 我国历年图书出版量

D. 我国历年黄金储备

E. 某地区国有企业历年资金利税率

4. 某企业2003—2012年的商品销售额见表5-27。

表5-27　某企业2003—2012年的商品销售额

年　　份	商品销售额/万元	年　　份	商品销售额/万元
2003	9.60	2008	14.00
2004	10.66	2009	15.04
2005	13.52	2010	18.28
2006	14.76	2011	17.96
2007	13.08	2012	18.70

则(　　)。

A. 此时间数列有逐渐上升的长期趋势

B. 可采用3项、4项或5项移动平均法

C. 5项移动平均比3项移动平均修匀程度更大些

D. 3项移动平均后的数列要比原数列项数减少2项

E. 用4项移动平均还需做一次“正位平均”



5. 某公司 8 月上旬各天职工人数见表 5－28。

表 5－28　某公司 8 月上旬各天职工人数

日　　期	1	2	3	4	5	6	7	8	9	10
出勤人数/人	300	304	298	302	301	296	298	302	300	299

则(　　)。

A. 这是总量指标时间数列

B. 这是时点指标时间数列

C. 该公司 8 月上旬平均每天出勤人数 300 人

D. 求该公司 8 月上旬平均每天出勤人数是求序时平均数

E. 求该公司 8 月上旬平均每天出勤人数采用简单算术平均法

6. 下列数列属于由一个时期数列和一个时点数列对比构成的相对数或平均数时间数列的是(　　)。

A. 历年职工平均工资时间数列

B. 工业企业全员劳动生产率数列

C. 历年职工人均消费支出时间数列

D. 历年恩格尔系数时间数列

E. 百元产值能耗时间数列



7. 用水平法计算平均发展速度的公式是(　　)。

A. $\dfrac{\sum af}{\sum f}\times 100\%$

B. $\sqrt[n]{\dfrac{a_n}{a_0}}\times 100\%$

C. $\sqrt[n]{\prod x}\times 100\%$

D. $\dfrac{\dfrac{a_1}{2}+a_2+a_3+\cdots+a_{n-1}+\dfrac{a_n}{2}}{n-1}\times 100\%$

E. $\sqrt[n]{R}\times 100\%$

8. 某地区粮食总产量见表 5－29。

表 5－29　某地区粮食总产量

年　　份	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
产量/万吨	230	236	241	246	252	257	262	276	281	286

则(　　)。

A. 该地区粮食生产发展趋势接近于直线型

B. 以 2007 年为 $t=-1$，以 2008 年为 $t=1$，直线趋势方程是 $\hat{Y}_t=a+bt=256.7+3.17t$

C. 时间每前进一年，产量平均增长 3.17 万吨

D. 若要利用 $\hat{Y}_t=a+bt=256.7+3.17t$ 预测 2013 年产量，则 t 应取 10

E. 利用 $\hat{Y}_t=a+bt=256.7+3.17t$ 预测 2013 年产量为 291.57 万吨



9. 2007—2011 年我国普通高等教育招生人数见表 5-30。

表 5-30 2007—2011 年我国普通高等教育招生人数

年　份	2007	2008	2009	2010	2011
普通高等教育招生人数/万人	566	608	640	662	682

则(　　)。

A. 2011 年比 2007 年普通高等教育招生人数增加了 116 万人

B. 2011 年比 2010 年普通高等教育招生人数增加了 20 万人

C. 2011 年普通高等教育招生人数是 2010 年的 1.03 倍

D. 2011 年普通高等教育招生人数是 2007 年的 1.2 倍

E. 2011 年普通高等教育招生人数比 2010 年增加了 3.02%

10. 下面的现象侧重于用几何平均法(水平法)计算平均发展速度的是(　　)。

A. 商品销售额　　B. 基本建设投资额

C. 垦荒造林数量　　D. 居民消费支出状况

E. 产品产量

三、判断题

1. 在各种时间数列中，指标值的大小都受到指标所反映的时期长短的制约。(　　)

2. 一个时间数列有 10 项，则其平均增长量是 10 个逐期增长量的简单算术平均。(　　)

3. 若逐期增长量每年相等，则其各年的环比发展速度是年年下降的。(　　)

4. 相邻的两个定基发展速度相除，得出相应的环比增长速度。(　　)

5. 根据两个时期数列对比所形成的相对指标时间数列来计算平均发展水平，实际上其平均发展水平等于两个时期数列指标和之比。(　　)

6. 呈直线趋势的时间数列，其各期环比发展速度大致相同。(　　)

7. 若时间数列中某项指标数值为 0，则不宜计算时间数列的平均发展速度。(　　)

8. 时间数列中，若某年环比增长速度为 0，则不能计算平均增长速度。(　　)

9. 某公司产品销售额在前 5 年环比增长速度分别为 5%、8%、11%、10%、11%，则平均增长速度为 9%。(　　)

10. 时间数列中若逐期增长量大致相等，则可以采用最小平方法配合直线方程或者指数曲线方程。(　　)

四、填空题

1. 发展水平可以表现为总量指标，也可以表现为________或________。

2. 某市 2013 年年初人口为 180 万人，比 2009 年年初增加 20 万人，则年平均增长量为________。

3. 各年末商品库存量数列属于____________数列，各年的基本建设投资额数列属于________数列。

4. 对1996—2012年的钢产量用移动平均法进行修匀。若想得到1998年的修匀数据，移动平均数最多为________年。

5. 静态平均数是根据________计算的，序时平均数则是根据________计算的。

6. 按照某城市十年发展规划，该市人均绿化面积要在2010年的人均4m^{2}的基础上10年后即2020年翻一番，则每年的平均增长速度是________。

7. 2010年第六次人口普查时我国大陆人口为133 972万人，以此为基数，若要求在2020年年底，人口数控制在15亿以内，则10年内我国人口增长率应为________。

8. 在移动平均法中，移动项数越________，则修匀数列越平滑。

9. 某公司生产某产品，1月生产2 000件，废品率为0.9%；2月生产2 200件，废品率为1.1%；3月生产1 900件，废品率为0.8%。则第一季度该公司产品的平均废品率为________。

10. 若没有季节变动，则季节比率应该为________。

五、应用能力训练题

1. 新华小区自行车库4月1日存自行车320辆，4月6日调出70辆，4月18日进货120辆，4月26日调出80辆，直至月末自行车数量未发生变动。问该库4月平均库存自行车多少辆?

2. 根据2006—2011年我国国民经济和社会发展统计公报资料(表5-31)。

表5-31 2006—2011年国民经济和社会发展统计公报

年　　份	2006	2007	2008	2009	2010	2011
国内生产总值比上年增长速度	—	14.2%	9.6%	9.2%	10.4%	9.2%

计算2006—2011年我国国内生产总值平均增长速度。

3. 我国2007—2011年年末国家外汇储备见表5-32。

表5-32 2007—2011年年末国家外汇储备

年　　份	2007	2008	2009	2010	2011
年末国家外汇储备/亿美元	15 282	19 460	23 992	28 473	31 811

计算2007—2011年年末我国平均国家外汇储备额。

4. 某商场第三季度各月商品资料见表5-33。

表5-33 某商品第三季度各月商品资料

月　　份	6月	7月	8月	9月
商品销售额/万元	141.6	144.0	148.2	154.7
月末商品库存额/万元	64.2	51.0	47.8	52.0

要求计算：

(1) 第三季度各月商品流转次数。

(2) 第三季度平均各月商品流转次数。



5. 某商店商品销售额及销售人员人数见表5-34。

表5-34 某商店商品销售额及销售人员人数

月　份	3	4	5	6
商品销售额/万元	165.0	198.0	177.0	216.9
月末销售人员人数/人	200	205	208	206

根据表5-34计算：

(1) 第二季度该店平均每月商品销售额。

(2) 第二季度平均售货员人数。

(3) 第二季度平均每个售货员的销售额。

(4) 4、5、6月各月(分别)的平均每个售货员的销售额。

(5) 第二季度平均每月每个售货员的销售额。

6. 某化工企业2008—2012年的化肥产量见表5-35。

表5-35 某化工企业2008—2012年的化肥产量

年　份	2008	2009	2010	2011	2012
化肥产量/万吨	400%			484%	
环比增长速度	—	5%			12.5%
定基发展速度	100%		111.3%		

根据表5-35计算：

(1) 表中空格所缺的统计指标。

(2) 平均增长量指标。

(3) 平均发展速度指标。

(4) 平均增长速度指标。

7. 已知2012年我国国内生产总值(GDP)为519 322亿元(初步核算数据)，若以平均每年增长7%的速度发展，试计算到2015年将达到什么水平。

8. 某地区2012年第三产业增加值为100亿元，如果以后每年增长8%，试问多少年后才能达到200亿元?

9. 请从时间长短、起伏规律和形成原因3个方面判断下面的现象属于时间数列构成因素中的哪一个。

(1) 超市的顾客人数在周末达到高峰。

(2) 银行的活期储蓄额在发放工资前减少，在发放工资后增多。

(3) 我国的进出口贸易额从长时间看是不断增长的。

(4) 由于“三聚氰胺”事件的曝光，部分奶制品需求量的急剧下降。

(5) 耐用消费品如电视、冰箱周期性更新导致需求量变化。

(6) 铁路部门运送旅客的客运量，在一年中有几个时段为高峰，另外几个时段为低谷。

(7) 玩具的销售量每年在12月达到最大额。

(8) 据中国互联网络中心(CNNIC)调查显示，我国网民上网高峰集中在20时和21时，最高峰是20时。有超过60%的网民经常在这一时点上网，次高峰在14时。低谷则是凌晨到早上8时。

10. 某企业2008—2012年资料见表5-36。

表5-36　某企业2008—2012年利润与成本费用

年　份	2008	2009	2010	2011	2012
利润总额/万元	250	320	350	500	700
成本费用总额/万元	2 000	2 500	3 000	2 800	3 500
成本费用利润率	12.5%	12.8%	11.67%	17.86%	20%

计算2008—2012年平均成本费用利润率。

11. 某游览点历年观光游客的资料见表5-37。

表5-37　某游览点2006—2012年观光游客数

年　份	时间(t)	游客/百人
2006	−3	100
2007	−2	112
2008	−1	125
2009	0	140
2010	1	155
2011	2	168
2012	3	180
合计	0	980

请按照题中所设定的时间取值，采用最小平方法配合直线趋势方程，并预测2013年的游客人数。

12. 某地区甲产品2009—2012年各季收购量统计资料见表5-38。

表5-38　某地区甲产品收购量统计资料

年　份	第一季度/万吨	第二季度/万吨	第三季度/万吨	第四季度/万吨
2009	15	7	10	20
2010	16	8	12	20
2011	18	10	14	24
2012	21	17	19	27



根据表 5－38 资料：

(1) 用移动平均法对该动态数列进行修匀。

(2) 用按月(季)平均法计算其季节比率。

(3) 假设 2013 年全年计划收购量为 120 万吨，则将计划目标分解，每季度计划收购量应分别为多少？

(4) 若 2013 年上半年收购量已达 75 万吨，则照此态势第三季度收购量可达多少？

13. 美国内华达职业健康诊所(Nevada Occupational Health Clinic)是一家私人医疗诊所，它位于内华达州的斯帕克斯(Sparks)市。这个诊所专攻工业医疗，并且在该地区经营已经超过 15 年。1991 年年初，该诊所进入了增长的阶段。在其后的 26 个月里，该诊所每个月的账单收入从 57 000 美元增长到超过 300 000 美元。直至 1993 年 4 月 6 日，当诊所的主建筑物被烧毁时，诊所一直经历着戏剧性的增长。

诊所的保险单包括实物财产和设备，也包括出于正常商业经营的中断而引起的收入损失。确定实物财产和设备在火灾中的损失额，受理财产的保险索赔要求是一个相对简单的事情。但是确定在进行重建诊所的 7 个月中，收入的损失额是很复杂的，它涉及业主和保险公司之间的讨价还价。对如果没有发生火灾，诊所的账单收入“将会有什么变化”的计算，没有预先制定的规则。为了估计损失的收入，诊所用一种预测方法测算在 7 个月的停业期间将要实现的营业增长。在火灾前的账单收入的实际历史资料为拥有线性趋势和季节成分的预测模型提供了基础资料。这个预测模型使诊所得到损失收入的一个准确的估计值，这个估计值最终被保险公司所接受。

这是一个时间数列分析方法在保险业务中的成功案例。受此案例启发，当你对企业销售收入和销售成本进行预测时，应该怎样操作？与同学一起讨论，并阐述你的思路。



第6章 统计指数

理论目标

(1) 理解统计指数的含义和种类。
(2) 掌握编制统计指数的基本原则和方法。
(3) 掌握因素分析基本思路和步骤。

能力目标

(1) 能识别实践中复杂现象综合变动的问题与统计指数的关系，并能正确计算。
(2) 能够运用指数体系理论对复杂社会经济现象进行因素分析。

做一做

KITTY公司主要从事跑鞋、服装和旅行包3种商品的出口贸易。表6-1是该公司2011年和2012年两年3种商品出口量和出口单价资料。请根据表中资料计算并回答，从公司的角度出发，出口的3种商品“总的说来”(或者是“平均说来”)2012年比2011年出口单价及出口量分别是上升还是下降？程度是多少？

表6-1 KITTY公司商品出口单价和出口量

商品名称	计量单位	出口单价/元		出口量/万	
		2011年	2012年	2011年	2012年
跑鞋	双	40	45	120	150
服装	套	65	58	109	140
旅行包	个	180	200	28	22
合计	—	—	—	—	—



想一想

根据表6-1中资料，单独看任何一种商品出口单价及出口量都可以知道2012年与2011年相比变动的方向和程度(利用普通的动态相对数，即广义的指数)。例如，反映跑鞋出口价格的变化，为45/40=112.5%，即2012年与2011年相比，跑鞋出口价格上升了12.5%；反映旅行包出口量的变化，为22/28=78.57%，即2012年与2011年相比，旅行包出口量下降了21.43%。但是若要反映公司总出口价格和总出口量的综合变动情况，能否将3种商品简单加总或者平均？显然不能。那如何反映此类不能直接加总的复杂现象的综合变动呢？这就需要引进一种新的统计工具——统计指数。

6.1 统计指数的概念和种类

6.1.1 统计指数的概念

统计指数(index number)简称指数，它是对社会经济现象进行数量变动分析的一种特殊方法。指数的编制最早起源于物价指数。早在1650年英国人赖斯·沃汉(Rice Voughan)首创物价指数，用于度量物价的变化情况。其后，指数的应用范围不断扩大，其含义也随之发生了变化。

从指数的含义上看，它有广义与狭义之分。广义指数是指任何两个数值对比形成的相对数。因此，前面章节中我们讨论过的计划完成程度相对数、比较相对数、动态相对数等都可以叫指数。狭义指数是指反映不能直接加总、对比的复杂经济现象总体综合变动的特殊的动态相对数。这种不能直接加总、对比的总体是指数理论着重研究的对象，称为复杂现象总体，反映复杂现象总体综合变动状况的指数称为总指数。本章所讨论的指数是狭义指数。

如上例中，实际上，我们需要求出的是一种特殊的2012年与2011年比较的动态相对数，它能够反映3种不能直接加总的商品(跑鞋、服装、旅行包)的综合变动，这个特殊的动态相对数就是我们所说的狭义指数。

可见，统计指数具有如下的特点：一是相对性，指数是一个相对数，它可以度量总体各变量在不同时间的相对变化；二是综合性，指数能够反映复杂现象的综合变动，这是就狭义指数而言的，也是指数理论和方法的核心问题，没有综合性，指数就不可能发展成为一种独立的理论和方法论体系；三是平均性，统计指数所反映的复杂现象的综合变动，实际上也是各简单现象变动的代表性数值，反映变动的一般水平，因而具有平均的含义。

正是因为统计指数所具有的特点，它在社会经济生活中起到了重要的作用。

首先，统计指数能反映复杂现象总体在数量上的变动方向和变动程度。统计指数一般是用百分比表示的相对指标，其比值大于或小于100%，表示上升或下降的变动方向；具体比100%大或小的数值，则说明升降变动的相对程度。

其次，统计指数能对现象的总体变动进行因素分析。包括现象总体总量指标、相对指标和平均指标受各个因素变动的影响程度。运用指数分析法，不仅可以从相对数方面分析各构成因素变动对现象总变动的影响程度，还可以从绝对数方面分析其影响的绝对额。

最后，利用连续编制的指数数列，还可以对复杂现象总体长时间发展变化趋势进行分

析。这种方法尤其适用于对比分析那些性质不同却有联系的时间数列的变动关系。因为，指数数列的对比能够克服不同性质现象不可比的问题。例如，编制零售物价指数数列，反映一定时期内物价的变动情况；编制每一天的股价指数，反映股票价格的升与跌；等等。

6.1.2 统计指数的种类

1. 个体指数和总指数

按指数所反映的对象范围的不同，统计指数可分为个体指数、组(或类)指数和总指数。个体指数是反映单个经济现象变动情况的相对数，属于广义指数。例如，个别产品的物量指数，个别商品的价格指数，等等。显然，个体指数是在简单现象总体的条件下存在的。常用的个体指数如下：

$$\text{商品销售量个体指数} \quad k_q = \frac{q_1}{q_0}$$

$$\text{价格个体指数} \quad k_p = \frac{p_1}{p_0}$$

$$\text{成本个体指数} \quad k_z = \frac{z_1}{z_0}$$

式中：q_1为报告期销售量；q_0为基期销售量；p_1为报告期价格；p_0为基期价格；z_1为报告期成本；z_0为基期成本；k 为个体指数。

总指数是反映复杂现象总体在数量上平均变动程度的相对数，用 $\bar{k}$ 表示。例如，综合反映多种产品产量平均变动程度的产品产量总指数 $\bar{k}_q$，综合反映多种产品成本平均变动程度的成本总指数 $\bar{k}_z$，等等。再如，工业产品总产值指数和零售商品物价总指数等总指数，都是在复杂现象总体的条件下进行编制的。它的计算形式有综合指数和平均指数。指数法的应用，要与科学分组相结合，因而在编制总指数的同时，往往还要编制组指数或类指数，借以反映个体内部各部分现象数量上的变动程度。组(或类)指数是反映某一组(或类)指标变动情况的相对数。它是相对于个体指数而言的，实质上还是总指数。

2. 数量指标指数和质量指标指数

按指数所表明的指标性质的不同，统计指数可分为数量指标指数(index number of quantity)和质量指标指数(index number of quality)。数量指标指数反映研究对象总体总规模的变化程度，如工业产品产量指数、商品销售量指数等；质量指标指数反映工作质量或成绩好坏、管理水平高低等变动情况，如产品价格指数、劳动生产率指数、产品成本指数等。在指数的应用中，必须重视这种数量指标指数与质量指标指数的区分，采用不同的编制方法，进行不同情况的动态分析。

3. 定基指数和环比指数

按指数采用基期的不同，统计指数可分为定基指数和环比指数。定基指数指各个时期指数都是采用同一固定时期为基期计算的；环比指数是依次以前一时期为基期计算的。定基指数的基期不依分析时期的变动而变化，可用来反映现象在一个较长时期的变动情况；



而环比指数的基期随报告期的变化而变化，可用来反映被研究现象逐期变动的情况。

4. 综合指数和平均数指数

按指数的编制方法不同，统计指数可分为综合指数和平均数指数。综合指数是由两个总量指标对比而形成的指数；平均数指数是以个体指数为基础，采用加权平均方法计算而成的指数。这两种指数形式既各具独立意义，又相互有联系。

5. 动态指数和静态指数

按指数所反映现象的时间不同，统计指数可分为动态指数和静态指数。动态指数是由两个不同时期的同一现象指标值对比而形成的指数；静态指数是指由同一时期不同空间同类现象指标值对比而形成的指数，以及同一空间范围内计划指标与实际指标对比而形成的指数。

6.2 综合指数与平均数指数

6.2.1 综合指数

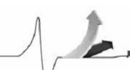
1. 综合指数的编制原则

总指数的编制有综合指数和平均数指数两类方法，综合指数是总指数计算的基本形式和出发点。下面以表6-2所示例子说明综合指数的编制要点。

表6-2 KITTY公司商品出口单价和出口量及其变动方向

商品名称	计量单位	出口单价/元		出口单价变动方向	出口量/万		销售量变动方向
		2011年	2012年		2011年	2012年	
跑鞋	双	40	45	↑上升	120	150	↑上升
服装	套	65	58	↓下降	109	140	↑上升
旅行包	个	180	200	↑上升	28	22	↓下降
合 计	—	—	—	—	—	—	—

这是一个简化和抽象了的仅由3种商品(实际中可能会有成千上万种商品)构成的复杂总体。按照指数的概念，可以得到商品出口量和出口单价两个方面的变动情况。例如，跑鞋，可以根据资料计算得其个体出口量指数为$k_q=150/120=125\%$；个体出口价格指数为$k_p=45/40=112.5\%$。而对反映3种商品出口量和出口价格总的变动情况的相对数即总指数，出于商品的不同属性，使得我们既不能直接将商品的出口量或单价加总起来对比计算总指数，也不能取其个体指数的简单平均数计算总指数。在这里，两个时期的各种商品出口量有增有减，出口单价有涨有跌，但所研究的这两个特征的变动在总量上又无法直接加总对比得出，这种不能直接加总对比的复杂现象，称为不同度量现象。综合指数的首要问题就是要使不能直接加总的各种商品的数量特征改变为能够加总，进而可以对比的现象总



量。商品是使用价值与价值的统一体，商品出口数量总是和出口价格联系在一起的。若将二者结合起来，就能把各种商品的不同使用价值形态变成相同形态的价值量，使得不同度量的总体量，变为可以加总即可以同度量的总体量。

由此引出综合指数编制的两个要点。第一个要点是确定同度量因素，使复杂总体中不能直接加总的量过渡到能直接加总的量。所谓同度量因素，是指在总指数计算的过程中，为解决总体的构成单位及其数量特征不能加总(即不能同度量)的问题，而使用的一个媒介因素或转化因素。它有两个作用：一是同度量作用；二是权数作用。第二个要点是需要将相应的同度量因素固定在某一个水平上。例如，为了反映各种商品销售数量总的变动情况，就需要将同度量因素——销售价格固定在某一水平(如基期价格、报告期价格或固定价格水平)上，以便进行对比。如果要反映各种商品销售价格总的变动情况，也需将对应的商品销售量固定在某一水平上。

针对同度量因素采用的时期不同，国际上比较通用的有两种方法：拉氏综合指数和派氏综合指数。拉氏综合指数由德国的经济统计学家拉斯贝尔(Laspeyres)1864 年首次提出，该方法主张将同度量因素固定在基期；派氏综合指数由德国的经济统计学家派许(Paasche)1874 年首先提出，该方法主张将同度量因素固定在报告期。

在统计实践中，为了有统一的计算口径，我们通用的统计指数编制原则是在编制数量指标指数时，以质量指标作为同度量因素，将它固定在基期；在编制质量指标指数时，以数量指标作为同度量因素，将它固定在报告期。

2. 数量指标指数

数量指标指数采用拉氏公式，其计算公式为

$$\bar{k}_q=\frac{\sum q_1 p_0}{\sum q_0 p_0}\times 100\% \tag{6-1}$$

根据表 6-2 所列的资料可以得到所需的对比指标值，见表 6-3。

利用公式(6-1)计算商品出口量指数，即

$$\bar{k}_q=\frac{\sum q_1 p_0}{\sum q_0 p_0}\times 100\%=\frac{19\ 060}{16\ 925}\times 100\%\approx 112.61\%$$

$$\sum q_1 p_0-\sum q_0 p_0=19\ 060-16\ 925=2\ 135(\text{万元})$$

表 6-3　商品出口量指数计算

商品名称	计量单位	出口单价(p)/元		出口量(q)/万		出口额(pq)/万元			
		2011 年 p_0	2012 年 p_1	2011 年 q_0	2012 年 q_1	2011 年实际 p_0q_0	2012 年实际 p_1q_1	假定出口额 p_1q_0	假定出口额 p_0q_1
跑鞋	双	40	45	120	150	4 800	6 750	5 400	6 000
服装	套	65	58	109	140	7 085	8 120	6 322	9 100
旅行包	个	180	200	28	22	5 040	4 400	5 600	3 960
合计	—	—	—	—	—	16 925	19 270	17 322	19 060



计算结果说明：

(1) 出口量综合变动的方向和程度，2012 年出口量比 2011 年增加了 12.61%。

(2) 出口量变动对出口额的影响程度，即变动使 2012 年商品出口额比 2011 年增加了 12.61%。

(3) 出口量变动对出口额的绝对影响量，即

$$\sum q_1 p_0 - \sum q_0 p_0 = 19\ 060 - 16\ 925 = 2\ 135(\text{万元})$$

对公司而言，是多出口了商品使出口额增加了 2 135 万元。

3. 质量指标指数

质量指标指数采用派氏公式，其计算公式为

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100\% \tag{6-2}$$

运用公式(6-2)，根据表 6-3 所列资料算出出口价格总的变动为

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100\% = \frac{19\ 270}{19\ 060} \times 100\% \approx 101.10\%$$

$$\sum p_1 q_1 - \sum p_0 q_1 = 19\ 270 - 19\ 060 = 210(\text{万元})$$

计算结果表明：在假定报告期出口量不变的情况下，报告期 3 种商品价格总水平比基期上升了 1.10%；由于价格总水平上升，使得出口额增加了 210 万元。

6.2.2 平均数指数

用综合指数法编制总指数，无论选择哪一种同度量因素，都是把不可度量的变量(如产量、价格)转化为可相加的变量(如产值、销售额)。如果研究的范围比较大，产品的种类比较多，要取得基期和报告期相对应的全面资料是很难的。这就需要从掌握的资料出发，采用总指数的另一种编制形式——平均数指数。平均数指数是从个体指数出发，先计算质量指标和数量指标的个体指数，然后采用加权平均的方法编制总指数。常用的有两种：加权算术平均数指数和加权调和平均数指数。

1. 加权算术平均数指数

加权算术平均数指数是对个体指数按加权算术平均法加以计算，即以个体指数为变量值，以一定时期的总值资料为权数，加权算术平均以计算总指数的方法。

【实例 6.1】 某大学生自主创业，经营绿色早餐摊点，主要销售馅饼、油条两种传统食品。销售资料见表 6-4。

求该摊点总的说来 8 月与 7 月比较销售量的变动情况。

解题步骤：

(1) 根据已知资料，标出相应字母符号。

表 6-4　绿色早餐摊点销售资料

商品名称	计量单位	实际销售额/元		8月销售量比7月销售量增长的百分比
		7月	8月	
馅饼	个	1 500	1 680	−6.67%
油条	千克	9 000	10 450	5.56%
合计	—	10 500	12 130	—

（2）先写出销售量的综合指数公式 $\bar{k}_q=\dfrac{\sum q_1p_0}{\sum q_0p_0}\times 100\%$。

（3）已知分母项资料 $\sum q_0p_0$，缺分子项资料。但根据个体指数 $k_q=\dfrac{q_1}{q_0}$，可推导出 $q_1=k_qq_0$，则公式变换为

$$\bar{k}_q=\frac{\sum k_qq_0p_0}{\sum q_0p_0}\times 100\% \tag{6-3}$$

这里 p_0q_0 为权数，因表现形式与加权算术平均数相同，因而取名加权算术平均数指数。

（4）根据公式中需要的内容，列出相应计算栏目，见表 6-5。

表 6-5　绿色早餐摊点销售资料计算

商品名称	计量单位	实际销售额/元		8月销售量比7月销售量增长的百分比	k_q	$k_qq_0p_0$
		7月(q_0p_0)	8月(q_1p_1)			
馅饼	个	1 500	1 680	−6.67%	0.933 3	1 400
油条	千克	9 000	10 450	5.56%	105.56	9 500
合计	—	10 500	12 130	—	—	10 900

（5）将数值带入公式，计算得

$$\bar{k}_q=\frac{\sum k_qq_0p_0}{\sum q_0p_0}\times 100\%\approx 103.81\%$$

$$\sum k_qq_0p_0-\sum q_0p_0=10\ 900-10\ 500=400(\text{元})$$

结果表明：该早餐摊点8月与7月比较，总销量上升了3.81%；由于销量的增加，使销售收入增加了400元。

加权算术平均数指数一般适用于数量指数。在实践中，$q_0p_0/\sum q_0p_0$ 常用 W 表示，W 代表权数，这样，即使没有各产品基期资料 q_0p_0，但具有产品价值总量结构 W 资料，就可以求出相应的指数。

2. 加权调和平均数指数

加权调和平均数指数是对个体指数按加权调和平均法加以计算，即以个体指数为变量



值，以一定时期的总值资料为权数，加权调和平均以计算总指数的方法。

【实例6.2】 某大学生自主创业，经营绿色早餐摊点，主要销售馅饼、油条两种传统食品。销售资料见表6-6。

表6-6 绿色早餐摊点销售情况

商品名称	计量单位	实际销售额/元		8月销售量比7月销售量增长的百分比
		7月	8月	
馅饼	个	1 500	1 680	20%
油条	千克	9 000	10 450	10%
合计	—	10 500	12 130	—

求该摊点总的说来8月与7月比较销售价格的变动情况。

解题步骤：

（1）根据已知资料，标出相应字母符号。

（2）先写出销售价格的综合指数公式 $\bar{k}_p=\dfrac{\sum p_1q_1}{\sum p_0q_1}\times100\%$。

（3）已知分子项资料 $\sum p_1q_1$，缺分母项资料。但根据个体指数 $k_p=\dfrac{p_1}{p_0}$，可推导出 $p_0=\dfrac{1}{k_p}p_1$，则公式变换为

$$\bar{k}_p=\frac{\sum p_1q_1}{\sum\frac{1}{k_p}p_1q_1}\times100\% \tag{6-4}$$

这里 p_1q_1 为权数，该公式因表现形式与加权调和平均数相同而得名。

（4）根据公式中需要的内容，列出相应计算栏目，见表6-7。

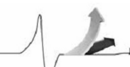
表6-7 绿色早餐摊点销售情况计算

商品名称	计量单位	实际销售额/元		8月销售量比7月销售量增长的百分比	k_p	$\frac{1}{k_p}p_1q_1$
		7月(q_0p_0)	8月(q_1p_1)			
馅饼	个	1 500	1 680	20%	1.2	1 400
油条	千克	9 000	10 450	10%	1.1	9 500
合计	—	10 500	12 130	—	—	10 900

（5）将数值带入公式，计算得

$$\bar{k}_p=\frac{\sum p_1q_1}{\sum p_0q_1}\times100\%=\frac{\sum p_1q_1}{\sum\frac{1}{k_p}p_1q_1}\times100\%\approx111.28\%$$

$$\sum p_1q_1-\sum\frac{1}{k_p}p_1q_1=12\ 130-10\ 900=1\ 230(\text{元})$$



结果表明：该早餐摊点8月与7月比较，销售价格上升了11.28%；由于价格的增加，使销售收入增加了1 230元。

加权调和平均数指数一般适用于质量指数。

平均数指数与综合指数既有区别，又有联系。二者的联系在于，在一定的权数下，平均数指数是综合指数的一种变形。但是，作为一种独立的指数形式，平均数指数在实际中不仅作为综合指数的变形使用，而且它本身也具有独特的应用价值。

6.3　指数体系及因素分析

通过前面对指数编制方法的介绍看到，一个指数通常只能说明某一方面的问题，因此在实践中往往需要将多个指数结合起来形成相应的指数体系，以便对相互联系的经济现象做深入分析。

6.3.1　指数体系的含义与作用

1. 指数体系的含义

社会经济现象的变动往往受多种因素的影响，各种因素不是孤立存在的，而是相互联系和相互影响的。例如，反映全国居民生活水平状况，可以用职工工资总额指数、人均可支配收入指数、人均纯收入指数、居民消费水平指数、居民消费价格指数等。这些指数从不同侧面反映居民收入、消费状况，具有一定的内在联系。可以说这些指数构成了反映居民生活状况的指数体系。又如，工业总产值是由产品产量和产品出厂价格两个因素构成，商品销售额是由商品销售量和销售价格两个因素构成的，产品利润是由生产量、产品价格和销售利润率构成的，等等。因而，工业总产值、商品销售额、产品利润等的变动，必然会受到各构成因素的影响。于是，我们将那些经济上有联系，在数量上保持一定关系的若干指数形成的整体，称作指数体系。上述例子，可用指数体系反映它们的联系，即

工业总产值指数＝产品产量指数×出厂价格指数

商品销售额指数＝商品销售量指数×物价指数

销售利润指数＝销售量指数×销售价格指数×销售利润率指数

指数体系包括两个方面的基本内容：①各因素指数的乘积应等于总变动指数；②各因素指数分子、分母差额总和应等于总量指标指数实际发生的总差额。

2. 指数体系的作用

指数体系的主要作用表现在以下两个方面。

(1) 进行指数间的推算。根据指数体系中各指数间的相互关系，当已掌握指数体系中若干个指数时，对剩下的其中某个未知指数，就可运用指数体系的关系进行推算。例如，已知2013年5月与上年同期相比，某地区物价综合上涨了7.7%，销售额没变，则销售量指数为100%÷107.7%≈92.85%，表明2013年与上年相比销售量下降了7.15%。

(2) 进行因素分析。因素分析是利用指数体系从数量方面分析现象总变动中，各影响



因素对其影响的方向、程度及绝对效果。

6.3.2 因素分析应用举例

因素分析是依据指数体系的理论，分析多因素影响的社会经济现象总变动中，各因素的影响方向和影响程度的方法。

进行因素分析一般有4个步骤。

第一步：分析被研究对象及其影响因素。这里的被研究对象是具体的统计指标，如商品销售额、流通费用额、原材料费用总额等。当明确了被研究现象是某个统计指标时，就要分析这个统计指标含有哪些影响因素，这是因素分析的基础。

第二步：建立指数体系。相对数关系式表现为现象总体指数等于各影响因素指数的乘积；绝对数关系式表现为现象总体指数分子与分母的差额等于各影响因素分子与分母差额之和。

第三步：搜集资料，计算指数体系两个关系式中各项数值。按公式内容和要求，搜集有关资料，并进行整理、计算。

第四步：根据计算结果，做出分析结论和简要的文字说明。

利用指数体系进行因素分析，依据分析对象不同可分为对总量指标、相对指标和平均指标变动进行的因素分析。

1. 总量指标变动的因素分析

总量指标变动的因素分析可按其影响因素的多少不同，分为两因素分析和多因素分析。

1）两因素分析

总量指标的两因素分析，有相对数和绝对额分析。

相对数分析：
$$\frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1} \tag{6-5}$$

绝对额分析：
$$\sum q_1 p_1 - \sum q_0 p_0 = \left(\sum q_1 p_0 - \sum q_0 p_0\right) + \left(\sum q_1 p_1 - \sum q_1 p_0\right) \tag{6-6}$$

【实例6.3】 以表6－8所列资料对销售额的变动进行因素分析。

表6－8 商品销售额变动的因素分析

商品名称	计量单位	价格(p)/元		销售量(q)		销售额(pq)/元		
		基期 (p_0)	报告期 (p_1)	基期 (q_0)	报告期 (q_1)	基期实际 (p_0q_0)	报告期实际 (p_1q_1)	假定销售额 (p_0q_1)
甲	米	25	20	400	600	10 000	12 000	15 000
乙	件	40	36	500	600	20 000	21 600	24 000
丙	千克	50	60	200	180	10 000	10 800	9 000
合计	—	—	—	—	—	40 000	44 400	48 000

第一步：列出分析对象，报告期与基期相比较销售额的变化。

$$\bar{k}_{qp}=\frac{\sum q_1p_1}{\sum q_0p_0}=\frac{44\ 400}{40\ 000}\times 100\%=111\%$$

$$\sum q_1p_1-\sum q_0p_0=44\ 400-40\ 000=4\ 400(\text{元})$$

第二步：建立指数体系。

$$\frac{\sum q_1p_1}{\sum q_0p_0}=\frac{\sum q_1p_0}{\sum q_0p_0}\times\frac{\sum p_1q_1}{\sum p_0q_1}$$

$$\sum q_1p_1-\sum q_0p_0=\left(\sum q_1p_0-\sum q_0p_0\right)+\left(\sum q_1p_1-\sum q_1p_0\right)$$

第三步：进行因素分析。

销售量总指数 $\bar{k}_q=\dfrac{\sum q_1p_0}{\sum q_0p_0}\times 100\%=\dfrac{48\ 000}{40\ 000}\times 100\%=120\%$

$$\sum q_1p_0-\sum q_0p_0=48\ 000-40\ 000=8\ 000(\text{元})$$

所以，由于销售量的变化，使销售额上升了20%，绝对额增加了8 000元。

销售价格总指数 $\bar{k}_p=\dfrac{\sum p_1q_1}{\sum p_0q_1}\times 100\%=\dfrac{44\ 400}{48\ 000}\times 100\%=92.5\%$

$$\sum q_1p_1-\sum q_1p_0=44\ 400-48\ 000=-3\ 600(\text{元})$$

所以，由于销售价格的变化，使销售额下降了7.5%，绝对额减少了3 600元。

第四步：综合影响。

$$111\%=120\%\times 92.5\%$$

$$4\ 400(\text{元})=8\ 000+(-3\ 600)$$

计算结果表明：销售额增长11%是销售量上升20%与价格下降7.5%共同作用的结果；销售额增加4 400元，是销售量上升使销售额增加8 000元、价格下降使销售额减少3 600元这两种因素共同作用的结果。

【实例6.4】 某商业企业有关资料见表6-9。

表6-9 某商业企业工资及职工人数资料

指标 \ 年份	2011	2012	指标 \ 年份	2011	2012
工资总额/万元	1 800	2 812.5	平均工资/（万元/人）	3.6	4.5
职工人数/人	500	625			

分析该企业工资总额的变动受职工人数和平均工资变动的影响情况，用 x 表示平均工资，f 表示职工人数，则

$$\text{工资总额指数}=\frac{\bar{x}_1f_1}{\bar{x}_0f_0}\times 100\%=\frac{2\ 812.5}{1\ 800}\times 100\%=156.25\%$$

$$\text{工资总额增加的绝对数额}=\bar{x}_1f_1-x_0f_0=2\ 812.5-1\ 800=1\ 012.5(\text{元})$$

$$\text{职工人数指数}=\frac{f_1}{f_0}\times 100\%=\frac{625}{500}\times 100\%=125\%$$



$$职工人数增长使工资总额增加的绝对额=(f_1-f_0)\bar{x}_0$$
$$=(625-500)\times 3.6=450(万元)$$

$$平均工资指数=\frac{\bar{x}_1}{\bar{x}_0}\times 100\%=\frac{4.5}{3.6}\times 100\%=125\%$$

$$职工人数增长使工资总额增加的绝对额=(\bar{x}_1-\bar{x}_0)f_1$$
$$=(4.5-3.6)\times 625=562.5(万元)$$

3 个指数之间的关系为

$$156.25\%=125\%\times 125\%$$
$$1\,012.5(万元)=450+562.5$$

计算结果表明：2012 年与 2011 年相比较，该企业工资总额增长了 56.25%，增加的绝对数额为 1 012.5 元。其中，职工人数增长了 25%，由于职工人数增长使工资总额的绝对数增加了 450 万元；平均工资提高了 25%，由于平均工资提高，使工资总额的绝对数额增加了 562.5 万元。

2）多因素分析

以上我们运用指数体系对总量指标的变动进行了两因素分析，这种分析方法还可以推广运用于 3 个或 3 个以上的因素分析。例如，工业产品原材料支出总额变动可分解为产量、单位产品原材料消耗量和单位原材料价格 3 个因素的变动影响。因此就需要编制原材料支出额指数及其所包括的 3 个因素指数的指数体系，来进行多因素变动的联系分析。又如，工业总产值动态变动可以分解为对职工人数、工人数占职工人数比重和工人劳动生产率 3 个因素进行变动影响分析，利税可以分解为销售量、销售价和利税率 3 个因素，等等。多因素的分析方法和两因素的分析方法基本原理是相同的，但由于多因素现象指数体系所包括的现象因素较多，指数的编制过程比较复杂。

因此，在进行多因素现象分析时，必须注意以下两点。

（1）在编制多因素指标所组成的综合指数时，为了测定某一因素指标的变动影响，要把两个或两个以上因素固定不变。这里仍然利用综合指数编制的一般原理，来确定固定因素所属时期，即在反映数量指标因素的变动影响时，应以基期的质量指标为同度量因素；而在反映质量指标因素的变动影响时，应以报告期数量指标为同度量因素。同时，由于所包括的因素较多，还要考虑多因素的合理排列顺序，来确定同度量因素。

（2）对综合指数中的多因素排列顺序，要具体分析现象总体的内容，依据现象因素的联系加以具体确定。现以工业产品原材料支出额的组成因素排列顺序为例加以说明，依据它们之间的联系，要按产量、单位产品原材料消耗量、单位原材料价格的顺序排列，即

$$原材料支出总额=\overbrace{产品产量\times\underbrace{\textbf{原材料单耗}}}^{原材料消耗量}\underbrace{\times 原材料单价}_{单位产品原材料消耗额}$$

这样，产品产量相对于单位产品原材料消耗额是数量指标，原材料单价相对于原材料消耗用量是质量指标，这与前面讲的两因素分析法统一起来。这样，就可以编制多个因素指数，用于分析这些因素变动对现象总体变动的影响作用。

下面以工业产品原材料支出额分解为 3 个因素的资料为例进行说明。设：m 为原材料单位消耗量，q、p 表示产量和原材料价格，则工业产品原材料支出额为 qmp，各因素指

数所形成的指数体系(相对数)为

$$\underbrace{\frac{\sum q_1 m_1 p_1}{\sum q_0 m_0 p_0}}_{\text{Ⅰ}} = \underbrace{\frac{\sum q_1 m_0 p_0}{\sum q_0 m_0 p_0}}_{\text{Ⅱ}} \times \underbrace{\frac{\sum q_1 m_1 p_0}{\sum q_1 m_0 p_0}}_{\text{Ⅲ}} \times \underbrace{\frac{\sum q_1 m_1 p_1}{\sum q_1 m_1 p_0}}_{\text{Ⅳ}} \tag{6-7}$$

其中：Ⅰ对应于原材料支出总额指数；Ⅱ对应于产品产量指数；Ⅲ对应于原材料单耗指数；Ⅳ对应于原材料单价指数。

$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = (\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0) + (\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0) + (\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0) \tag{6-8}$$

【实例 6.5】 以表 6-10 中所列资料为例，针对某企业生产 3 种产品的主要原材料的支出额及其分解因素资料，编制指数体系并进行多因素分析。

表 6-10 产品原材料费用额变动的因素分析

商品名称	计量单位	产量(q)		单耗(m)		原材料单价/元		原材料费用额/元			
		q_0	q_1	m_0	m_1	p_0	p_1	$q_0m_0p_0$	$q_1m_0p_0$	$q_1m_1p_0$	$q_1m_1p_1$
甲	个	600	900	1.0	0.9	15	18	9 000	13 500	12 150	14 580
乙	件	1500	1800	2.2	1.9	18	20	59 400	71 280	61 560	68 400
丙	只	900	900	2.5	2.0	25	24	56 250	56 250	45 000	43 200
合计	—	—	—	—	—	—	—	124 650	141 030	118 710	126 180

解：

(1) 原材料费用总额指数：

$$\bar{k}_{qmp} = \frac{\sum q_1 m_1 p_1}{\sum q_0 m_0 p_0} \times 100\% = \frac{126\ 180}{124\ 650} \approx 101.23\%$$

$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = 126\ 180 - 124\ 650 = 1\ 530(\text{元})$$

计算结果说明该企业报告期生产的 3 种产品的 3 种原材料的费用额比基期增长了 1.23%，绝对数额增加了 1 530 元。

(2) 分析各因素对总变动的影响情况，分别计算 3 个因素指数。

① 产量总指数：

$$\bar{k}_q = \frac{\sum q_1 m_0 p_0}{\sum q_0 m_0 p_0} \times 100\% = \frac{141\ 030}{124\ 650} \times 100\% \approx 113.14\%$$

$$\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0 = 141\ 030 - 124\ 650 = 16\ 380(\text{元})$$

计算结果说明由于报告期产量比基期增加，使原材料总费用额上升了 13.14%，其绝对数额增加了 16 380 元。

② 原材料单位消耗总指数：

$$\bar{k}_m = \frac{\sum q_1 m_1 p_0}{\sum q_1 m_0 p_0} \times 100\% = \frac{118\ 710}{141\ 030} \times 100\% \approx 84.17\%$$



$$\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0 = 118\ 710 - 141\ 030 = -22\ 320(\text{元})$$

计算结果说明了报告期产品原材料的单位消耗比基期下降了15.83%，由于单耗减少，使原材料费用额减少了22 320元。

③ 原材料价格总指数：

$$\bar{k}_p = \frac{\sum q_1 m_1 p_1}{\sum q_1 m_1 p_0} \times 100\% = \frac{126\ 180}{118\ 710} \times 100\% \approx 106.29\%$$

$$\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0 = 126\ 180 - 118\ 710 = 7\ 470(\text{元})$$

计算结果说明由于报告期原材料价格上涨，使原材料费用支出额增加了6.29%，增加支出费用为7 470元。

以上4种指数关系，满足 $\bar{k}_{qmp} = \bar{k}_q \times \bar{k}_m \times \bar{k}_p$

$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = \left(\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0\right) + \left(\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0\right) + \left(\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0\right)$$

亦即

$$101.23\% \approx 113.14\% \times 81.17\% \times 106.29\%$$

$$1\ 530(\text{元}) = 16\ 380 + (-22\ 320) + 7\ 470$$

2. 相对指标变动的因素分析

相对数是由两个指标对比得到的，这两个指标的变动都会影响相对数的变动，故亦可进行因素分析。因为 $C=A/B$（设 A 为数量指标，B 为质量指标），所以

$$\frac{C_1}{C_0} = \frac{\frac{A_1}{B_1}}{\frac{A_0}{B_0}} = \frac{\frac{A_1}{B_0}}{\frac{A_0}{B_0}} \times \frac{\frac{A_1}{B_1}}{\frac{A_1}{B_0}} \tag{6-9}$$

$$A_1/B_1 - A_0/B_0 = (A_1/B_0 - A_0/B_0) + (A_1/B_1 - A_1/B_0) \tag{6-10}$$

【实例6.6】 以表6-11中所列资料为例，针对某公司商品销售和库存资料，编制相对指标指数体系并进行因素分析。

表6-11 某公司商品销售和库存资料

指　　标	符　　号	2013年第一季度	2012年第一季度	指　　数
商品销售额/万元	A	A_1　3 609.00	A_0　2 755.00	131.00%
平均库存额/万元	B	B_1　1 432.00	B_0　1 450.00	98.76%
商品周转次数/次	C	C_1　2.52	C_0　1.90	132.63%

根据表中资料计算：$A_1/B_1 \approx 2.52$(次)；$A_0/B_0 = 1.90$(次)；

$$A_1/B_0 = 3\ 609/1\ 450 \approx 2.49(\text{次})$$

所以，

$$\frac{2.52}{1.90} = \frac{2.49}{1.90} \times \frac{2.52}{2.49}$$

$$132.63\% \approx 131.05\% \times 101.20\%$$

$$2.52 - 1.90 = (2.49 - 1.90) + (2.52 - 2.49)$$

$$0.62(\text{次})=0.59+0.03$$

计算结果表明：商品周转次数2013年第一季度与上年同季度比较增长了32.63%，这是因为销售额变动使之增长31.05%，平均库存额变动使之增长1.20%；商品周转次数增加了0.62次，是因销售额变动增加0.59次和因平均库存额变动增加0.03次两因素共同作用的结果。

3. 平均指标变动的因素分析

统计指数法还可用于平均指标的动态分析。平均指标指数就是用来反映总平均数变动及其原因的相对数，即同一经济内容两个不同时期的平均指标数值之比。它是对平均指标变动情况进行分析的一种方法。平均指标指数的一般公式为

$$k=\frac{\bar{x}_1}{\bar{x}_0} \tag{6-11}$$

式中：$\bar{x}_1$ 为报告期平均指标数值；$\bar{x}_0$ 为基期平均指标数值。

常见的平均指标指数有劳动生产率指数、平均工资指数、平均单位成本指数等。下面以劳动生产率指数为例，进一步介绍平均指标指数的含义及基本原理。

劳动生产率指数的公式为

$$k=\frac{\bar{x}_1}{\bar{x}_0}=\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}=\frac{\sum x_1\cdot\dfrac{f_1}{\sum f_1}}{\sum x_0\cdot\dfrac{f_0}{\sum f_0}} \tag{6-12}$$

式中：$\sum f_0$和$\sum f_1$分别代表基期和报告期的工人总人数；$\sum x_0 f_0$和$\sum x_1 f_1$分别代表基期和报告期的产量；x_0和x_1分别代表基期和报告期的劳动生产率，$\bar{x}_0$ 和 $\bar{x}_1$ 分别代表基期和报告期的平均劳动生产率；f 或 $\dfrac{f}{\sum f}$ 是一种结构相对数，简称结构。

由此可见，劳动生产率指数反映两个因素变动的影响，即各组工人劳动生产率变动的影响和各组工人人数在全部工人总数中所占比例变动的影响。这是因为加权算术平均数的大小取决于变量标志值和权数。因此，平均指标指数所反映的变动程度，也包括两个因素间的影响。为了测定一个因素的变动情况，必须将另一个因素固定下来。

1）平均指标指数编制

平均指标指数分析需要编制3种平均指标指数，它们是可变构成指数、固定构成指数和结构变动影响指数，并形成如下的指数体系：

可变构成指数＝固定构成指数×结构变动影响指数

它们分别简称为可变指数、固定指数和结构指数。下面分别进行介绍。

（1）可变指数。可变指数是报告期平均指标与基期平均指标之比。例如，在进行总平均工资变动的因素分析时，首先要计算总平均工资指数，计算总平均工资指数的报告期和基期总平均工资，是分别以各个时期工人数为权数对各组工资水平进行平均计算的。所以两期总平均工资的变动，不仅反映了各组(类)工资水平的变动，而且受各组工人数结构变



动的影响。这种包括结构变动影响作用的总平均工资指数，就是平均工资可变构成指数，其公式为

$$\text{平均工资可变构成指数 } k_{\text{可变}}=\frac{\bar{x}_1}{\bar{x}_0}=\frac{\sum x_1 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_0}{\sum f_0} \tag{6-13}$$

（2）固定指数。固定指数是假定构成不变$\left(\text{即 }\frac{f}{\sum f}\text{ 频率不变}\right)$，纯粹反映组平均数总变动的相对数。例如，在分析企业总平均工资变动中各组工资水平变动影响关系时，依据综合指数编制的原理，为了消除结构因素的变动影响，反映各组工资水平的变动程度，要把工人数加以固定，并且固定在报告期上。这种工人数结构固定的总平均工资指数，称为平均工资固定指数，其公式为

$$\text{平均工资固定指数 } k_{\text{固定}}=\frac{\sum x_1 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_1}{\sum f_1} \tag{6-14}$$

这一指数的经济内容：分子指标是报告期企业实际平均工资，分母指标是假定各组工资水平保持不变的情况下，企业报告期平均工资。它们之间的差别，只是由于两期各组工人工资水平变动所引起的。

（3）结构指数。结构指数是指假定组平均数不变(即 x_0 不变)，纯粹反映结构变动程度的相对数。例如，分析工人变动影响指数，必须将各组工人工资水平因素固定下来，并将它固定在基期水平上。其计算公式为

$$\text{平均工资结构变动影响指数 } k_{\text{结构}}=\frac{\sum x_0 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_0}{\sum f_0} \tag{6-15}$$

这个指数的分子指标是假定各组工人工资水平保持不变的情况下，企业报告期的平均工资，分母指标是基期企业实际平均工资。它们的对比关系，可以表明两期中各级工人数构成变动对企业平均工资变动的影响。

2）平均指标指数体系

通过以上分析可以看出，平均指标变动的因素分析，实质上是现象结构的变动分析，因为分析所使用的指数都与结构有关。也就是说，可变指数是包含了结构变动因素的平均指标指数，固定指数是排除了结构变动影响的平均指标指数，而结构指数是纯粹考虑了结构变动影响的平均指标指数。上述各种指数之间具有一定的内在联系，形成一个指数体系，即

相对数为

$$\frac{\sum x_1 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_0}{\sum f_0}=\left(\frac{\sum x_0 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_0}{\sum f_0}\right)\times\left(\frac{\sum x_1 f_1}{\sum f_1}\Bigg/\frac{\sum x_0 f_1}{\sum f_1}\right) \tag{6-16}$$

绝对数为

$$\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}=\left(\frac{\sum x_0 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}\right)+\left(\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_1}{\sum f_1}\right) \tag{6-17}$$

通过平均指标指数体系，用具体例子说明平均指标指数的分析方法及其应用。

【实例 6.7】 根据表 6－12 中的某企业工人人数、工资额资料，说明平均指标指数的分析及应用。

表 6－12　某企业职工人数和工资水平资料

职工类别	工人数/人		平均工资/元		工资总额/元		
	基期(f_0)	报告期(f_1)	基期(x_0)	报告期(x_1)	基期(x_0 f_0)	报告期(x_1 f_1)	假定(x_0 f_1)
老职工	520	320	3 720	4 200	1 934 400	1 344 000	1 190 400
新职工	470	720	1 920	2 200	902 400	1 584 000	1 382 400
合计	990	1 040	—	—	2 836 800	2 928 000	2 572 800

从资料中可以看出：该企业新老职工的平均工资报告期比基期都提高了，其中老职工人均提高 480 元，新职工人均提高 280 元。但全厂职工的平均工资是否也提高呢？下面通过计算来分析说明该企业总平均工资的变动情况及原因，根据表中资料求得

$$\bar{x}_0=\frac{\sum x_0 f_0}{\sum f_0}=\frac{2\ 836\ 800}{990}\approx 2\ 865.45(\text{元})$$

$$\bar{x}_1=\frac{\sum x_1 f_1}{\sum f_1}=\frac{2\ 928\ 000}{1\ 040}\approx 2\ 815.38(\text{元})$$

$$k_{\text{可变}}=\frac{\bar{x}_1}{\bar{x}_0}\times 100\%=\frac{\sum x_1 f_1}{\sum f_1}\bigg/\frac{\sum x_0 f_0}{\sum f_0}\times 100\%$$

$$=\frac{2\ 928\ 000}{1\ 040}\bigg/\frac{2\ 836\ 800}{990}\times 100\%\approx 98.25\%$$

$$\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}=2\ 815.38-2\ 865.45=-50.07(\text{元})$$

这说明该企业工人总平均工资报告期比基期减少了 1.75%，减少的金额为 50.07 元。

$$k_{\text{固定}}=\frac{\sum x_1 f_1}{\sum f_1}\bigg/\frac{\sum x_0 f_1}{\sum f_1}=\frac{2\ 928\ 000}{1\ 040}\div\frac{2\ 572\ 800}{1\ 040}\approx 113.81\%$$

$$\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_1}{\sum f_1}=2\ 815.38-2\ 473.85=341.53(\text{元})$$

这说明该企业工人组平均工资水平报告期比基期总的上升 13.81%，由此使总平均工资增加 341.53 元。

$$k_{\text{结构}}=\frac{\sum x_0 f_1}{\sum f_1}\bigg/\frac{\sum x_0 f_0}{\sum f_0}\times 100\%=2\ 473.85\div 2\ 865.45\times 100\%\approx 86.33\%$$

$$\frac{\sum x_0 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}=2\ 473.85-2\ 865.45=-391.6(\text{元})$$



计算结果表明：由于该企业工人人数结构的变动使总平均工资减少了13.67%，减少金额平均为每个职工391.6元。

现将上面分析计算结果用平均指标指数体系表示出来，为

$$\frac{\sum x_1 f_1}{\sum f_1} \bigg/ \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} \bigg/ \frac{\sum x_0 f_0}{\sum f_0}\right) \times \left(\frac{\sum x_1 f_1}{\sum f_1} \bigg/ \frac{\sum x_0 f_1}{\sum f_1}\right)$$

$$98.25\% \approx 113.81\% \times 86.33\%$$

$$\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0}\right) + \left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1}\right)$$

$$-50.07(\text{元}) = 341.53 + (-391.6)$$

综上，该企业全体工人平均工资报告期为基期的98.25%，比基期减少了1.75%，这是由于各组工人工资水平上升使总平均工资提高了13.81%，由于工人人数结构的变化使平均工资减少了13.67 %，两者共同作用使平均工资提高1.75%。每个工人的平均工资减少额为50.07元，是由于各组工资水平上升使其增加341.53元，工人结构变化使其减少了391.6元这两个因素共同作用所致。

6.4 常用价格指数简介

6.4.1 消费者价格指数

消费者价格指数(consumer price index)，英文缩写为CPI，我国称之为居民消费价格指数，它是反映城乡居民家庭购买并用于日常生活消费的一篮子商品和服务项目价格水平随时间而变动的相对数，一般以百分比为表达形式。CPI不是商品价格，是一组商品和服务项目价格变动的平均综合指标。CPI是在居民消费价格统计的基础上编制计算出来的。国际上公认CPI有三大用途，即通货膨胀指标、国民经济核算及支付的调整。

从2001年起，我国采用国际通用做法，逐月编制并公布以2000年价格水平为基期的居民消费价格定基指数，作为反映我国通货膨胀(或紧缩)程度的主要指标。经国务院批准，国家统计局城调总队负责全国居民消费价格指数的编制及相关工作，并组织、指导和管理各省区市的消费价格调查统计工作。

我国编制价格指数的商品和服务项目，根据全国城乡13.3万户城乡居民家庭消费支出构成资料和有关规定确定，目前共包括食品、烟酒、衣着、家庭设备用品及维修服务、医疗保健和个人用品、交通和通信、娱乐教育文化用品及服务、居住八大类，262个基本分类。

居民消费价格指数就是在对参加全国数据汇总的约500个样本市县包括超市、菜市场、百货商场、医院、旅行社等6.3万个采价点进行价格调查的基础上，根据国际规范的流程和公式算出来的。在美国构成该指标的主要商品共分七大类，其中包括食品、酒和饮品住宅，衣着，交通，医药健康，娱乐，其他商品及服务。在美国，消费物价指数由劳工统计局每月公布，有两种不同的消费物价指数。一是工人和职员的消费物价指数，简称

CPW；二是城市消费者的消费物价指数，简称 CPIU。

编制居民消费价格指数需要两类资料：一是各项消费商品和服务的价格变动数据；二是各项消费商品和服务在居民生活中的开支权重。此比例一般每 5 年更换一次，5 年期内各年度适当调整。

编制 CPI 简单地说，可以概括为三大步骤。

第一步：确定基本分类中每个指数商品的代表规格品。目前计算 CPI 的代表规格品既包括商品消费，也包括服务消费；既包括传统消费项目，也包括汽车等“现代消费”。大部分基本分类选择两三个规格品，目前我国编制指数的地区根据其规模选择了 550～750 个代表规格品。这些代表规格品的选择是根据近 4 万户城市居民家庭和 6 万多户农村居民的消费支出构成资料和有关规定决定的。一般而言，大城市的代表规格品多些，小城镇就少些。

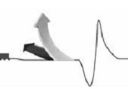
规格品的选取并不是随意确定的，选取 CPI 调查的代表规格品要遵循一定的原则。首先，要选择在当地消费、零售量大的商品；其次，要选取价格变动趋势和变动程度有较强的代表性，即选中规格品与未选中规格品的价格变动特征愈相关愈好；还要选择同一基本分类的规格品之间，性质差异愈大愈好，价格变动特征的相关性愈低愈好；对选中的规格品应具有较好的销售前景，工业消费品必须是合格产品，产品包装上应有注册商标、产地、规格等级等标识，不合格品不能选为调查代表规格品；每一个基本分类的代表规格品的选择数量，原则上不能少于制度规定的最低标准，可根据当地的实际情况适当增减；代表规格品一经确定，原则上一年内不能更改。为保证代表规格品的唯一性，选择代表规格品时要详细描述代表规格品的名称、品牌、产地、规格等级、货号等特征。需要说明的是，对失去代表性的商品和服务项目每年要进行一次调整、补充。

第二步：采集代表规格品的价格。价格调查采用定人、定点、定时的“三定”原则。也就是要求全国 4 000 余名现场调查员必须在固定的时间段内，对调查采价点内确定的调查商品或服务项目价格进行直接调查，调查人员在一段时间内必须是固定的，不能随意更换。对价格变动不频繁的商品和服务项目，如汽车、邮资、物管、报刊、医疗等每月采价 1 次；对于烟酒、家电、服装等一般性调查商品和服务项目，每月要去商场超市采价两三次；与居民生活密切相关、价格变动比较频繁的主副食品价格，如蔬菜、肉、禽、水果，每 5 天采集一次价格，即每个月采价 6 次。而且采集的价格必须是成交价，对于鲜活商品不能取早上的高价和晚上的低价，必须选择固定时间采集的价格，以便准确反映各种商品价格变动的情况。调查员收集的基础数据由县级调查队通过网络上报省级调查总队，省级调查总队审核后上报国家统计局。目前，在 50 个地级市已经实现由调查员通过手持数据采集器将基础数据向国家统计局直接报送，现在已在全国所有 CPI 调查市县推广使用。

第三步：进行价格指数的计算。价格指数的计算包括以下 3 个步骤：一是代表规格品平均价格的计算；二是基本分类指数的计算，包括月环比指数的计算等；三是类别及总指数逐级加权平均计算。

世界各国的消费价格指数多用加权算术平均法计算，我国也采用加权算术平均法。其公式为

$$\bar{k}_p = \sum k_{pW} \tag{6-18}$$



式中：$\bar{k}_p$ 代表总(类)指数；k_p 代表个体(或类)指数；W 代表权数。

权数是按总指数、大类、中类、小类和代表品分层计算的，每层的权数总和为100。各省的权数按中选的样本市县的资料计算，全国的权数则根据各省的资料计算。

计算时，先计算商品或服务项目的个体指数，然后计算小类、中类、大类指数，最后计算总指数。由小到大，逐步升级，层层平均。

现以城市居民消费价格指数为例，计算详见表6-13。

表6-13 某城市居民消费价格指数

(　　年　月/　　年　月)

城市编码：

商品类别及品名	规格等级牌号	计量单位	平均价格/元		权数	以上年为基期	
			上年	本年		指　数	指数×权数
(甲)	(乙)	(丙)	(1)	(2)	(3)	(4)	(5)=(4)×(3)
总指数					100	105.55%	—
一、食品类					34	114.21%	3 883.14
1. 粮食					20	110.40%	2 208
(1) 细粮					88	110.52%	9 725.76
面粉	普通粉	千克	4.16	4.74	38	113.94%	4 329.72
大米	标　二	千克	4.58	4.97	55	108.52%	5 968.6
江米	标　二	千克	5.20	5.59	5	107.50%	537.5
挂面	富强粉	千克	4.56	4.93	2	108.11%	216.22
(2) 粗粮					12	109.53%	1 314.36
2. 淀粉及薯类					1	111.02%	111.02
3. 干豆类及豆制品					2	119.06%	238.12
4. 油脂类					7	115.76%	810.32
5. 肉食及其制品					23	118.32%	2 721.36
6. 蛋类					6	109.87%	659.22
7. 水产品类					7	106.24%	743.68
8. 菜类					9	121.11%	1 089.99
9. 调味品					2	108. 98%	217.96
10. 糖类					2	102.45%	204.9
11. 茶及饮料					4	108.43%	433.72
12. 干鲜瓜果					4	112.08%	448.32
13. 糕点饼干					4	115.46%	461.84
14. 液体乳及乳制品					4	118.57%	474.28

续表

商品类别及品名	规格等级牌号	计量单位	平均价格/元		权数	以上年为基期	
			上年	本年		指数	指数×权数
15. 在外用膳食品					3	120.98%	362.94
16. 其他食品					2	117.89%	235.78
二、烟酒类					6	101.27%	607.62
三、衣着类					9	101.14%	910.26
四、家庭设备用品及维修服务					10	101.23%	1 012.3
五、医疗保健和个人用品					10	100.19%	1 001.9
六、交通和通信					11	99.78%	1 097.58
七、娱乐教育文化用品和服务					7	102.21%	715.47
八、居住类					13	102.05%	1 326.65

CPI 是宏观经济分析与决策，以及国民经济核算的重要指标。一般来说，CPI 的高低直接影响着国家的宏观经济调控措施的出台与力度，如央行是否调息、是否调整存款准备金率等。同时，CPI 的高低也间接影响资本市场(如股票市场)的变化。

CPI 月度数据由国家统计局通过新闻发布的形式统一对外发布，每月一次。公布形式包括国务院统一安排的新闻发布会和中华人民共和国国家统计局官方网站(http://www.stats.gov.cn)的传播。国家统计局发布 CPI 的时间，每年都有日程表。例如，2013 年居民消费价格指数月度报告，一般在每月 9 日 9 时 30 分发布。

消费价格指数公布内容：①全国及各省(区、市)CPI；②城市 CPI；③农村 CPI；④8 个大类类别指数的环比、同比数据。

6.4.2 股票价格指数

股票价格指数即股票指数，是由证券交易所或金融服务机构编制的表明股票行市变动的一种供参考的指示数字。它是通过一定方法计算出的衡量整个股票市场的股票价格总体水平和市场交易状况的一个指标。它是帮助投资者掌握股市现状和分析、判断股市变动趋势的非常重要的尺度和信号，是反映国家或地区社会经济和政治形势变动情况的最敏感指标，是整个经济的晴雨表和政治风云录。

股票价格指数因涨跌迅速，一般要求按日编制。它是以某年某月某日股价作为基期股价(基准日股价)，基准日指数通常定为 100，以后各日股价同基日股价计算出百分数，即为各日股票指数。

股票指数通常运用综合指数形式，一般以股票发行量为权数，也有以成交量为权数的。举一个简单例子说明股票指数编制原理，见表 6－14。

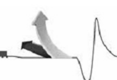


表6-14 股票资料

股票名称	发行量/股	基准日价格(p_0)/元	计算日股票价格/(元/股)		
			p_1	p_2	p_3
甲	20 000	15	22	19	18
乙	30 000	11	18	19	17
丙	40 000	9	7	10	9

则第一日股票指数为

$$\frac{\sum qp_1}{\sum qp_0}=\frac{22\times20\ 000+18\times30\ 000+7\times40\ 000}{15\times20\ 000+11\times30\ 000+9\times40\ 000}\times100\%$$

$$=\frac{1\ 260\ 000}{990\ 000}\times100\%\approx127.27\%$$

第二日股票指数为

$$\frac{\sum qp_2}{\sum qp_0}=\frac{19\times20\ 000+19\times30\ 000+10\times40\ 000}{15\times20\ 000+11\times30\ 000+9\times40\ 000}\times100\%$$

$$=\frac{1\ 350\ 000}{990\ 000}\times100\%\approx136.36\%$$

第三日股票指数为

$$\frac{\sum qp_3}{\sum qp_0}=\frac{18\times20\ 000+17\times30\ 000+9\times40\ 000}{15\times20\ 000+11\times30\ 000+9\times40\ 000}\times100\%$$

$$=\frac{1\ 230\ 000}{990\ 000}\times100\%\approx124.24\%$$

这说明，第一日股价上涨27.27%，第二日比第一日上涨了9.09个百分点，第三日又比第二日下跌了12.12个百分点。

6.4.3 几种常见的股票指数

1. 道·琼斯股票指数

道·琼斯股票指数即道·琼斯股票价格平均指数，是世界上最有影响、使用最广的股价指数。它以在纽约证券交易所挂牌上市的一部分有代表性的公司股票作为编制对象，由4种股价平均指数构成：①以30家著名的工业公司股票为编制对象的道·琼斯工业股价平均指数；②以20家著名的交通运输业公司股票为编制对象的道·琼斯运输业股价平均指数；③以6家著名的公用事业公司股票为编制对象的道·琼斯公用事业股价平均指数；④以上述3种股价平均指数所涉及的65家公司股票为编制对象的道·琼斯股价综合平均指数。在4种道·琼斯股价指数中，以道·琼斯工业股价平均指数最为著名，它被大众传媒广泛地报道，并作为道·琼斯指数的代表加以引用。

道·琼斯股票指数是世界上历史最为悠久的股票指数。第一次公布道·琼斯指数是在

1884 年 7 月 3 日，由道·琼斯公司的创始人之一、《华尔街日报》首任编辑查尔斯·亨利·道(Charles Henry Dow)编制，登载在其属下的《华尔街日报》上。最初是根据 11 种具有代表性的铁路公司的股票，采用算术平均法进行计算编制而成。自 1897 年起，道·琼斯股票价格平均指数开始分成工业与运输业两大类，其中工业股票价格平均指数包括 12 种股票，运输业平均指数则包括 20 种股票，并且开始在道·琼斯公司出版的《华尔街日报》上公布。在 1929 年，道·琼斯股票价格平均指数又增加了公用事业类股票，使其所包含的股票达到 65 种，并一直延续至今。

现在的道·琼斯股票价格平均指数是以 1928 年 10 月 1 日为基期，因为这一天收盘时的道·琼斯股票价格平均数恰好约为 100 美元，所以就将其定为基准日。而以后股票价格同基期相比计算出的百分数，就成为各期的股票价格指数，所以现在的股票指数普遍用点做单位，而股票指数每一点的涨跌就是相对于基准日的涨跌百分数。

道·琼斯股票价格平均指数是目前世界上影响最大、最有权威性的一种股票价格指数，在纽约证券交易所营业时间里，每隔半小时公布一次。这一股票价格平均指数自编制以来从未间断，可以用来比较不同时期的股票行情和经济发展情况，成为反映美国股市行情变化最敏感的股票价格平均指数之一，是观察市场动态和从事股票投资的主要参考。当然，由于道·琼斯股票价格指数是一种成分股指数，它包括的公司仅占目前 2500 多家上市公司的极少部分，而且多是热门股票，且未将近年来发展迅速的服务性行业和金融业的公司包括在内，所以它的代表性也一直受到人们的质疑和批评。

2. 标准·普尔股票价格指数

除了道·琼斯股票价格指数外，标准·普尔股票价格指数在美国也很有影响，它是美国最大的证券研究机构即标准·普尔公司于 1923 年开始编制发表股票价格指数。最初采选了 230 种股票，编制两种股票价格指数。到 1957 年，这一股票价格指数的范围扩大到 500 种股票，分成 95 种组合。其中最重要的 4 种组合是工业股票组、铁路股票组、公用事业股票组和 500 种股票混合组。从 1976 年 7 月 1 日开始，改为 400 种工业股票，20 种运输业股票，40 种公用事业股票和 40 种金融业股票。几十年来，虽然有股票更迭，但始终保持为 500 种。

3. 日经道·琼斯股价指数

日经道·琼斯股价指数(日经平均股价)系由日本经济新闻社编制并公布的反映日本股票市场价格变动的股票价格平均数。该指数从 1950 年 9 月开始编制。最初根据东京证券交易所第一市场上市的 225 家公司的股票算出修正平均股价，当时称为“东证修正平均股价”。1975 年 5 月 1 日，日本经济新闻社向道·琼斯公司买进商标，采用美国道·琼斯公司的修正法计算，这种股票指数也就改称“日经道·琼斯平均股价”。1985 年 5 月 1 日在合同期满 10 年时，经两家商议，将名称改为“日经平均股价”。

按计算对象的采样数目不同，该指数分为两种。一种是日经 225 种平均股价。其所选样本均为在东京证券交易所第一市场上市的股票，样本选定后原则上不再更改。由于日经 225 种平均股价从 1950 年一直延续下来，因而其连续性及可比性较好，成为考察和分析日



本股票市场长期演变及动态的最常用和最可靠的指标。该指数的另一种是日经500种平均股价。这是从1982年1月4日起开始编制的。由于其采样包括500种股票，其代表性就相对更为广泛，但它的样本是不固定的，每年4月要根据上市公司的经营状况、成交量和成交金额及市价总值等因素对样本进行更换。

4. 恒生指数

恒生指数(Hang Seng Index；Hang Seng，HSI)，由香港恒生银行全资附属的恒生指数公司编制。它是香港股票市场上历史最久、影响最大的股票价格指数，由香港恒生银行于1969年11月24日开始发表。指数是由若干只成分股(即蓝筹股)市值计算出来的，代表了香港交易所所有上市公司的12个月平均市值涵盖率的70%，恒生指数由恒生银行属下恒生指数有限公司负责计算及按季检讨，公布成分股调整。恒生指数最初从香港500多家上市公司中挑选出来的33家有代表性且经济实力雄厚的大公司股票作为成分股，分为四大类——4种金融业股票、6种公用事业股票、9种地产业股票和14种其他工商业(包括航空和酒店)股票。2006年6月30日宣布将成分股数目由33只加至38只，后来经过多次增减变动，2012年5月10日，第一次将博彩业公司“金沙中国”加入成分股，2012年11月17日，“昆仑能源”成为成分股。至此成分股增加到50只。目前对恒生指数影响最大的成分股是汇丰控股(15%)，其次是中国建设银行(6.72%)、中国移动(6.68%)、中国工商银行(6.08%)。

恒生股票价格指数的编制是以1964年7月31日为基期，基点确定为100点。其计算方法是将成分股股票按每天的收盘价乘以各自的发行股数为计算日的市值，再与基期的市值相比较，乘以100就得出当天的股票价格指数。开始时，恒生指数只计算一个总指数，为方便投资者获得更多信息，自1984年1月13日起，增算金融业、地产业、工商业和公用事业4个分类指数。由于恒生股票价格指数所选择的基期适当，因此，不论股票市场狂升或猛跌，还是处于正常交易水平，恒生股票价格指数基本上能反映整个股市的活动情况。

恒生指数历史最高点数为2007年10月30日，突破31 900点，升至31 958.41点，历史最低收市点数则是1967年8月31日的58.61点。

5. 上证指数

上证指数是由上海证券交易所编制并发布的指数系列，包括上证综合指数、上证180指数、上证50指数、上证A股指数、上证B股指数、上证分类指数、上证债券指数和上证基金指数等。其中最早编制的为上证综合指数。因此，狭义的上证指数指的是“上证综指”。

上证综指(Shanghai Composite Index)，全称“上海证券交易所综合股价指数”，是国内外普遍采用的反映上海股市总体走势的统计指标。它以上海证券交易所挂牌上市的全部股票为计算范围，以发行量为权数综合计算，于1991年7月15日公开发布。上证指数以“点”为单位，基日定为1990年12月19日，基日指数定为100点。

随着上市股票品种逐渐增加，上海证券交易所在综合指数的基础上增加了其他指数。1992年2月21日，增设上证A股指数与上证B股指数，以反映不同股票(A股、B股)的各自走势。1993年6月1日，又增设了上证分类指数，即工业类指数、商业类指数、地产

业类指数、公用事业类指数、综合业类指数，以反映不同行业股票的各自走势。

上证指数是一个派许公式计算的以报告期发行股数为权数的加权综合股价指数。计算公式为

$$报告期指数=(报告期采样股的市价总值/基日采样股的市价总值)\times 100$$

$$市价总值=\sum(市价\times 发行股数)$$

其中，基日采样股的市价总值亦称为除数。

上证180指数(又称上证成分指数)是上海证券交易所对原上证30指数进行了调整并更名而成的，其样本股是在所有A股股票中抽取最具市场代表性的180种样本股票，自2002年7月1日起正式发布。作为上证指数系列核心的上证180指数的编制方案，目的在于建立一个反映上海证券市场的概貌和运行状况、具有可操作性和投资性、能够作为投资评价尺度及金融衍生产品基础的基准指数。

上证180指数与通常计算的上证综指之间最大的区别在于，它是成分指数，而不是综合指数。成分指数是根据科学客观的选样方法挑选出的样本股形成的指数，所以能更准确地认识和评价市场。而综合指数包含了市场上所有的股票，在反映市场状况上就存在不少缺陷。例如，目前上证综指采用全市场平均市盈率标准，将不少业绩差、规模小、股价过高的股票包含进来，导致较高的市盈率。据测算，上证180指数目前的市盈率约28倍，比上证综指38倍的市盈率降低了约10倍。

上海证券交易所股票指数的发布几乎是和股市行情的变化同步，投资者据此可以检验自己投资的效果，并用以预测股票市场的动向。它是我国股民和证券从业人员研判股票价格变化趋势必不可少的参考依据。同时，新闻界、公司老板乃至政界领导人等也以此为参考指标，来观察、预测社会政治及经济发展形势。

6. 深证指数

深证指数是指由深圳证券交易所编制的股价指数，该股票指数的计算方法基本与上证指数相同，其样本为所有在深圳证券交易所挂牌上市的股票，权数为股票的总股本。由于以所有挂牌的上市公司为样本，其代表性非常广泛，且它与深圳股市的行情同步发布，它是股民和证券从业人员研判深圳股市股票价格变化趋势必不可少的参考依据。

现深圳证券交易所并存着两个股票指数，一个是老指数——深圳综合指数，另一个是现在的深圳成分股指数。但从运行势态看，两个指数间的区别并不是特别明显。深证综合指数以1991年4月3日为基日，1991年4月4日开始发布，基日指数为100。成分股指数以1994年7月20日为基日，1995年1月23日开始发布，基日指数定为1 000。

深证综合指数将在深圳证券交易所上市的全部股票纳入指数计算范围，当有新股上市时，在其上市后第二天纳入成分股，当某一股票暂停买卖时，则将其剔除于计算之外。若采样的股本结构有所变动，则改用变动之日为新基期，并以新基数计算。同时用连锁的方法将计算得到的指数追溯原有基期日，以维持指数的连续性。深证综合指数采用派氏加权价格指数，即以指数股的计算日股份数作为权数进行加权计算。股份数等于全部上市公司的总股本数。计算公式为

$$即日指数=(即日指数股总市值/基日指数股总市值)\times 基日指数$$



深圳成分股指数根据从上市公司中挑选出来的40家成分股进行计算。计算方法和计算公式与深证综合指数一致，只是其股份数等于成分股的可流通股本数。

本章小结

本章由浅入深地介绍了统计指数这个工具。首先需要掌握统计指数的概念和种类，然后能够根据资料进行计算，最后能够正确运用指数体系进行因素分析。

本章统计指数是指反映不能直接加总、对比的复杂社会经济现象总体综合变动的特殊动态相对数。具有综合性、相对性、平均性的性质，具有多种类型。

统计指数的计算，根据资料及方法不同，有综合指数和平均数指数两种算法。

综合指数的编制有两种，数量指标指数和质量指标指数。在实际应用中，一般地，数量指标指数以基期的质量指标作为同度量因素，即选用拉氏公式；质量指标指数以报告期的数量指标作为同度量因素，即选用派氏公式。

数量指标指数 $\bar{k}_q = \dfrac{\sum q_1 p_0}{\sum q_0 p_0}$　　质量指标指数 $\bar{k}_p = \dfrac{\sum p_1 q}{\sum p_0 q}$

平均数指数的计算形式有两种：算术平均数指数和调和平均数指数。

算术平均数指数 $\bar{k}_q = \dfrac{\sum k_q q_0 p_0}{\sum q_0 p_0}$　　调和平均数指数 $\bar{k}_q = \dfrac{\sum q_1 p_1}{\sum \frac{1}{k_q} q_1 p_1}$

统计指数这个工具的应用，通常是通过指数体系体现的。指数体系是若干个有联系的指数构成的系统。它有两个对等关系，即各影响因素指数的乘积等于现象总体指数；各影响因素指数变动额之和等于现象总体变动额。指数体系是进行因素分析的基本依据；利用指数体系，可以进行指数间的相互推算。

因素分析有总量指标、相对指标和平均指标因素分析。在进行总量指标因素分析时，应注意：为了分析某一因素的变动影响，需要将其他所有因素固定不变；并且依据因素之间的关系，对各因素的衔接按照数量指标在前、质量指标在后的顺序进行排列。对平均指标进行因素分析，将指数体系分解为3个相互关联的指数，即可变构成指数、固定构成指数和结构变动影响指数。

应用与拓展

2011年地区发展与民生指数报告

为客观、全面反映各地区发展与民生状况，中国统计学会组织有关专家研究并测算了各地区发展与民生指数(Development and Life Index，DLI)。该指数是在原综合发展指数的基础上，为了更加突出地反映民生改善情况，调整补充部分指标后修订而成的。地区发展与民生指数评价指标体系包括经济发展、民生改善、社会发展、生态建设、科技创新、公众评价六大方面，共42项指标。指数的计算与合成借鉴了联合国人类发展指数(HDI)等有关方法，根据指标的上、下限阈值来计算各个指标的评价指数(即无量纲化)，指数介于0～100，根据指标权重合成分类指数和总指数。测算结果显示，2011年

各地区发展与民生指数均比上年有所提高。

1. 四大区域发展与民生指数稳步提高，中、西部地区增速较快

2011年，四大区域的发展与民生指数，东部地区最高，为69.53%，比上年提高2.50个百分点；东北地区次之，为60.22%，比上年提高2.19个百分点；中部地区和西部地区分别为58.33%和55.41%，分别比上年提高2.73和2.79个百分点。从指数的增速看，西部地区最快，为5.30%；中部地区次之，为4.91%；东北地区和东部地区分别为3.78%和3.73%。

测算结果还表明，从2000年以来，我国四大区域发展与民生指数均稳步提高。2011年与2000年相比，东部地区由46.66%提高到69.53%，中部地区由36.98%提高到58.33%，西部地区由34.61%提高到55.41%，东北地区由40.55%提高到60.22%。

2. 四大区域分项指数均有不同程度的提高

发展与民生指数由经济发展、民生改善、社会发展、生态建设和科技创新5个类别的分项指数组成，与2010年相比，2011年四大区域的各分项指数均有不同程度的提高。

东、中、西部和东北地区经济发展类指数分别为80.38%、64.32%、64.24%和74.71%，比上年分别提高1.24、2.22、3.01、1.69个百分点。

东、中、西部和东北地区民生改善类指数分别为73.89%、63.42%、57.96%和65.05%，比上年分别提高3.20、2.75、3.56、3.10个百分点。

东、中、西部和东北地区社会发展类指数分别为69.15%、66.71%、64.66%和66.99%，比上年分别提高3.23、3.30、3.58、1.78个百分点。

东、中、西部和东北地区生态建设类指数分别为70.98%、62.93%、58.86%和59.82%，比上年分别提高1.92、2.19、1.02、3.25个百分点。

东、中、西部和东北地区科技创新类指数分别为41.10%、18.12%、15.60%和15.63%，比上年分别提高2.84、3.40、2.04、0.44个百分点。

3. 31个省(区、市)发展与民生指数普遍提高，差距有所缩小

2011年，31个省(区、市)的发展与民生指数比上年均有不同程度提高。指数排在前十名的地区分别为北京、上海、天津、江苏、浙江、广东、福建、山东、辽宁和重庆，指数均在60%以上。指数增速排在前十名的地区分别为甘肃、重庆、青海、河南、海南、安徽、贵州、新疆、四川和云南，均达到5%以上。

全国31个省(区、市)发展与民生指数的变异系数，自2004年后呈现逐年缩小的趋势，表明省际差距在缩小。2011年，变异系数又由上年的0.172缩小到0.162。

4. 各地区发展与民生指数和人均GDP的比较

从2011年各地区发展与民生指数和人均GDP的比较看，差异较大。有些省份人均GDP排序相对靠前，但发展与民生指数排序相对靠后，如内蒙古自治区、宁夏回族自治区、新疆维吾尔自治区、青海省等；一些省份人均GDP排序虽然相对靠后，但发展与民生指数排序相对靠前，如四川、江西、海南、安徽等省。

(资料来源：http://www.stats.gov.cn/tjfx/jdfx/t20130208_402872712.htm.)



习题与实训

一、单项选择题

1. 从广义指数的角度看，下面不属于指数的是(　　)。



A. 比较相对指标　　B. 居民人均生活费收入
C. 职工平均工资　　D. 计划完成程度相对指标

2. 下列指数中属于质量指标指数的是(　　)。
A. 产品产量指数　　B. 商品出口量指数
C. 股票价格指数　　D. 商品销售量指数

3. 总指数的两种计算基本形式是(　　)。
A. 个体指数和综合指数
B. 算术平均数指数和调和平均数指数
C. 综合指数和平均数指数
D. 可变构成指数、固定构成指数和结构变动影响指数

4. 如果消费价格指数上涨20%，则现在1元(　　)。
A. 只值原来的0.8元　　B. 只值原来的0.83元
C. 与原来的1元等值　　D. 无法与原来比较

5. 价格下降后，花同样多的钱可多购买基期商品的10%，则物价指数为(　　)。
A. 90%　　B. 90.9%　　C. 110%　　D. 111.1%

6. 某公司生产甲乙丙3种产品，上月产值分别为20 000元、30 000元、40 000元，本月产量与上月相比分别上升了10%、20%、5%，那么3种产品产量平均比上月上升(　　)。
A. 11.67%　　B. 35%　　C. 11.1%　　D. 无法计算

7. 某公司销售甲乙丙3种商品，本月销售额分别为23 760元、26 400元、5 016元，本月价格分别是上月的90%、100%、95%，那么3种商品价格与上月相比(　　)。
A. 平均上升5%　　B. 平均下降5%
C. 平均上升3.8%　　D. 平均下降3.8%

8. 已知某商场的商品销售量指数为105%，由于销售量增长而增加的销售额为10万元；又知道销售价格指数为110%，由于价格上涨而增加的销售额应为(　　)。
A. 30万元　　B. 21万元　　C. 20万元　　D. 18万元

9. 某商店今年与去年相比，商品销售量指数下降了10%，销售价格指数上涨10%，则商品销售额将(　　)。
A. 不变　　B. 上升
C. 下降　　D. 可能上升也可能下降

10. 某企业报告期产量比基期增长了10%，生产费用增长了8%，则其产品单位成本降低了(　　)。
A. 2%　　B. 20%　　C. 1.8%　　D. 18%

二、多项选择题

1. 某商品在2013年第一季度售出100千克，第二季度售出120千克，指数为120%，该指数是(　　)。
A. 综合指数　　B. 总指数　　C. 个体指数
D. 数量指标指数　　E. 质量指标指数

2. 编制价格综合指数，同度量因素可选择(　　)。

A. 成本　　B. 产量　　C. 销售量

D. 劳动生产率　　E. 出口量

3. 某地区商业企业职工去年全员劳动生产率指数为 132%，这是(　　)。

A. 个体指数　　B. 总指数

C. 平均指标指数　　D. 数量指标指数

E. 质量指标指数

4. 平均指标变动因素分析的指数体系中包括的指数有(　　)。

A. 可变构成指数　　B. 加权算术平均数指数

C. 固定构成指数　　D. 结构变动影响指数

E. 加权调和平均数指数

5. 某企业基期产值为 100 万元，报告期产值比基期增长 14%，又知以基期价格计算的报告期假定产值为 112 万元，则经过计算可知(　　)。

A. 由于产量变化使产值增加 12 万元

B. 产量增加 12%

C. 由于产量变化使产值增加 20 万元

D. 价格增加 12%

E. 由于价格变化使产值增加 2 万元

6. 消费者价格指数(CPI)是(　　)。

A. 数量指数　　B. 质量指数

C. 加权算术平均数指数　　D. 加权调和平均数指数

E. 静态指数

7. 下列属于数量指标指数的有(　　)。

A. 工业生产指数　　B. 劳动生产率指数

C. 职工人数指数　　D. 产品产量指数

E. 产品单位成本指数

8. 进行 3 个或 3 个以上多因素分析时，下面正确的是(　　)。

A. 为了测定某一因素指标的变动影响，要把两个或两个以上因素固定不变

B. 要对综合指数中的多因素按顺序排列

C. 因素排序时一般数量因素在前，质量因素在后

D. 因素排序时一般质量因素在前，数量因素在后

E. 因素分析方法和步骤与二因素一致

9. 在平均指标指数体系中(　　)。

A. 所谓可变指数，是指报告期平均指标与基期平均指标之比

B. 可变指数等于固定指数与结构指数乘积

C. 固定指数是假定构成不变$\left(\text{即}\dfrac{f}{\sum f}\text{不变}\right)$，纯粹反映组平均数总变动的相对数

D. 结构指数是指假定组平均数不变(即 x_0 不变)，纯粹反映结构变动程度的相对数

E. 对平均指标指数的分析是二因素分析



10. 某地区市场销售额 2012 年为 50 万元，比 2011 年增加了 10 万元，销售量与 2011 年相比上升了 5%，那么(　　)。

A. 市场销售量总指数是 105%

B. 市场销售价格总指数是 118.95%

C. 市场销售价格总指数是 119.05%

D. 由于销售量的变动使销售额增加 8 万元

E. 由于销售量的变动使销售额增加 2 万元

三、判断题

1. 某公司若员工工资水平今年与去年相比较全部上升，则总平均工资一定不会下降。(　　)

2. 运用拉氏公式计算统计指数时，须将同度量因素固定在报告期；同理，运用派氏公式计算统计指数时，须将同度量因素固定在基期。(　　)

3. 分析复杂现象总体的数量变动时，若研究的是数量指标的变动，则选择的同度量因素是数量指标。(　　)

4. 某企业的某种产品单位成本 2013 年 3 月与去年同期相比上升了 8%，产量下降了 8%，则总成本没升也没降。(　　)

5. 可变构成指数=固定构成指数÷结构影响变动指数。(　　)

6. 设 p 表示价格，q 表示销售量，则 $\sum p_0q_1 - \sum p_0q_0$ 表示由于商品价格的变动对商品总销售额的影响。(　　)

7. 消费者价格指数(CPI)是质量指标指数，我国的消费价格指数采用的是加权算术平均法计算的。(　　)

8. 总指数有两种计算形式，即个体指数和综合指数。(　　)

9. 商品价格上涨 9%，商品销售量增长 2%，则商品销售额增长 18%。(　　)

10. 反映全厂职工的总平均工资的可变指数是 98.25%，则可断定报告期与基期相比较，全厂职工的平均工资水平下降，但每个职工的工资水平不一定下降。(　　)

四、填空题

1. 指数体系中，总量指数等于各因素指数________，总量指数分子与分母差额等于各因素指数分子与分母差额________。

2. 某商店商品销售额报告期和基期相同，报告期商品价格比基期提高了 10%，那么，报告期商品销售量比基期________。

3. 2012 年与 2011 年相比，某机关的职工平均工资提高了 5%，职工人数减少了 1%，则职工工资总额________。

4. 指数体系的主要作用表现在两个方面。一是________，二是________。

5. 在编制数量指标指数时，一般以________作为同度量因素，把它固定在________；在编制质量指标指数时，一般以________作为同度量因素，把它固定在________。

6. 平均指数是________的加权平均数。常用的基本形式有两种，一是加权________平均数指数，二是加权________平均数指数。

7. 单位成本指数$=\dfrac{\sum z_1q_1}{\sum z_0q_1}\times 100\%=97\%$，$\sum z_1q_1-\sum z_0q_1=-2\ 200$(元)，则相对数的含义是________，绝对数的含义是________。

8. 某商品价格发生变化，现在的100元只值原来的90元，则价格指数为________。

9. 某工厂各组工人工资比基期都提高了10%，然而全厂工人平均工资却降低2%，其原因是工人结构影响指数________的结果。

10. 某地区社会商品零售额2012年为9.89亿元，比2011年增加1.29亿元，零售物价指数涨了3%，则2012年商品销售量是2011年销售量的________。

五、应用能力训练题

1. 区分表6-15中各指数的种类。

表6-15　区分指数种类训练题

统计指数	个体指数	总指数	数量指标指数	质量指标指数
某一种商品单位成本指数				
20种商品的价格指数				
企业职工的平均工资指数				
某型号手机的销售量指数				
全国消费品零售价格指数				

2. 某商场3种商品的实际销售额资料见表6-16。

表6-16　商品实际销售额资料

商品名称	计量单位	实际销售额/万元		8月销售量比7月销售量增长的百分比
		7月	8月	
甲	件	1 000	1 200	5%
乙	袋	2 000	2 000	0%
丙	台	3 000	3 200	10%

计算分析下列问题：

(1) 各种商品的个体价格指数和个体销售量指数。

(2) 3种商品的物价总指数和销售量总指数。

(3) 3种商品的销售额总指数。

(4) 从相对数和绝对数上分析销售额总指数与价格总指数、销售量总指数之间的经济关系。

3. 某公司2013年第一季度和第二季度的产品产值和价格资料见表6-17。



表 6-17 第一、第二季度的产品产值和价格资料

产品名称	产品产值/万元		第二季度价格是第一季度价格的百分比
	第一季度	第二季度	
甲	400	600	105%
乙	550	450	115%
丙	200	350	98%
合计	1 150	1 400	—

要求：分析产品单价和产量的变动对产值的变动影响。

4. 某商店商品的价格和销售量资料见表 6-18。

表 6-18 商品价格和销售量资料

商品名称	计量单位	价格/元		销售量	
		基期	报告期	基期	报告期
甲	双	48	45	300	400
乙	件	30	32	400	500

要求：

(1) 计算某商店几种商品的价格总指数和销售量总指数。

(2) 指出在总销售额增长绝对值中，有多少绝对数是受价格因素影响的，有多少绝对数是受销售量因素影响的？

5. 某企业某种产品生产总成本 2012 年为 12.9 万元，比 2011 年多 9 000 元，单位产品成本比 2011 年降低 3%，试确定：

(1) 生产总成本指数。

(2) 产品产量指数。

(3) 由于单位成本降低而节约的绝对额。

6. 某地区 2012 年社会商品零售额为 2 570 亿元，比 2011 年增长 9.4%，剔除零售物价上升的因素，社会商品零售实际增长 7.3%，试计算：

(1) 2012 年与 2011 年比较，零售物价上升多少？

(2) 计算分析 2012 年比 2011 年社会商品零售额增长的数量和原因。

7. 某地区粮食作物的生产情况见表 6-19。

表 6-19 某地区粮食作物的生产情况

粮食作物	播种面积/亩		亩产/斤	
	2010 年	2012 年	2010 年	2012 年
水 稻	400	300	800	900
小 麦	600	900	400	420

请计算：

(1) 该地区粮食平均亩产量指数。

(2) 由于播种面积的变动对总平均亩产量变动的影响。

8. 某工业企业有资料见表6-20。

表6-20 某工业企业相关资料

产品名称	计量单位	单位成本/元		产量	
		2011年	2012年	2011年	2012年
甲	千克	100	100	100	115
乙	台	50	55	200	220
丙	件	20	25	300	315
合计	—	—		—	—

请计算：

(1) 3种产品的产量总指数和单位成本总指数。

(2) 由于产量的变动对总成本的影响。

(3) 由于单位成本的变动对总成本的影响。

9. 某管理局所属3个工厂生产同种产品，它们的单位产品成本和产量资料见表6-21。

表6-21 3个工厂的单位产品成本和产量资料

厂 名	产量/件		每件成本/元	
	2011年	2012年	2011年	2012年
一厂	1 600	2 400	10.0	9.0
二厂	1 800	2 400	10.4	9.2
三厂	2 400	1 600	9.6	9.6

根据上述资料：

(1) 分别计算3个工厂生产各种产品的2011年和2012年总平均成本。

(2) 计算平均成本指数，分析由于平均成本下降而节约的总成本金额。

(3) 在平均成本的总变动中，分析各工厂成本水平变动及各工厂产量结构变动的影响程度相对数和绝对值。

10. 中财集团下属甲乙两企业第三季度与第四季度的有关资料见表6-22。

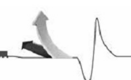


表6-22 甲乙企业第三、第四季度相关资料

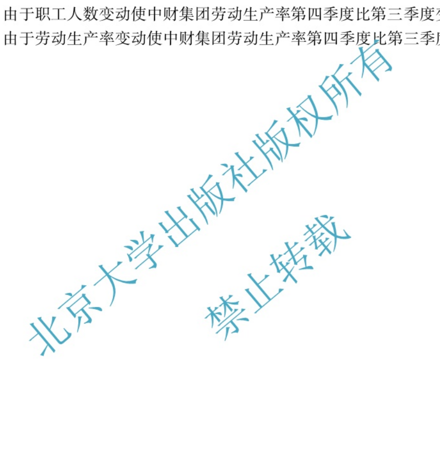
企业名称	平均职工人数/人		劳动生产率/元	
	第三季度	第四季度	第三季度	第四季度
甲	900	600	1 588	2 000
乙	1 100	1 400	2 909	3 429

请计算：

（1）中财集团第四季度与第三季度相比劳动生产率指数。

（2）由于职工人数变动使中财集团劳动生产率第四季度比第三季度变动程度。

（3）由于劳动生产率变动使中财集团劳动生产率第四季度比第三季度变动程度。





第7章 抽样推断

理论目标

(1) 理解抽样推断的基本概念。
(2) 掌握简单随机抽样形式的区间估计方法和必要样本单位数的确定方法。
(3) 理解假设检验的步骤。
(4) 掌握正态总体平均数和成数的假设检验方法。

能力目标

(1) 能根据研究目的选取抽样方式和方法、计算必要样本容量，进行抽样。
(2) 能根据样本资料采用区间估计方法对总体特征进行科学推断。
(3) 能对关于正态总体的平均数和成数问题进行假设检验。

做一做

张毅是大学学生会生活部的部长，近期生活部要对学生在学校食堂午餐花费情况进行调查。张毅所在的校区是一分校区，有学生 5 000 人左右，若对每位学生逐一调查既耗费时间又耗费人力，后来在老师的帮助下，他们采用抽样方法抽取了 100 名学生，得到午餐花费数据见表 7-1。

根据这些资料，如果以 95%把握推测，张毅所在分校区学生平均午餐花费为多少？同样以 95%可靠程度推断，分校区午餐花费在 4 元以下的学生人数大约有多少？



表7-1 学生在食堂午餐花费情况

午餐花费/元	人数/人	午餐花费/元	人数/人
4以下	8	8～10	10
4～5	22	10以上	7
5～6	38	合计	100
6～8	15		

想一想

张毅和他的同学为什么采用抽样调查方法？他们怎样计算出必需的样本单位数？采用什么方式去抽？而对已经抽出来的样本资料应该怎样科学推断全部总体的特征指标？这些问题的解决需要用到一个非常实用的统计工具——抽样推断，本章将简要介绍有关抽样推断的理论和方法。

7.1 抽样推断的基本概念

抽样推断(sampling inference)是在概率抽样调查(sampling survey)的基础上，利用样本的实际资料计算样本指标，并据此推算总体相应数量特征的一种统计分析方法。在前面的章节中，我们已经学习了计算各项综合指标，如总量指标、相对指标、平均数指标等反映总体的数量特征。但是，在许多场合我们不可能对总体的所有单位进行全面调查，从而达到对总体数量特征的认识。例如，城乡居民的电视收视率、市场商品需求量、民意测验等，由于时间、经费等原因，很难对每个单位进行观察，而只能组织抽样调查，取得部分的实际资料，来估计和判断总体的数量特征，以达到对现象总体的认识。

归纳起来，抽样推断具有4个特点。第一，抽样推断是由部分推算整体的一种认识方法。在第2章中已讲到，抽样调查是一种非全面调查方式，其调查目的并不是了解部分单位的情况，而是作为认识总体的一种手段。利用样本的部分信息，推断总体的相关信息。第二，抽样推断建立在随机取样的基础上。抽样调查可以是概率抽样也可以是非概率抽样，但是作为抽样推断基础的必须是概率抽样，即按随机原则抽取样本单位，这是抽样推断的前提。这样才能保证样本变量是随机变量，抽出的样本才具有代表性。第三，抽样推断是运用概率估计的方法，利用样本指标来估计总体参数。在数学上运用不确定的概率估计法，而不是运用确定的自变量和因变量的关系，它不能利用一定的函数关系来推算总体参数。原则上把由样本观察值所决定的样本指标看作随机变量。因此，用样本指标值推断总体指标值，是以一定可靠性做保证的。第四，抽样推断的误差是可以预先控制的，用样本指标值推断总体指标值是存在一定误差的。但是，由于抽样推断是有科学根据的，所以，抽样误差的范围可以事先通过有关资料加以计算，并且加以控制，从而使抽样推断具有一定的可靠程度，这是其他估算方法所不及的。下面介绍有关抽样的基本概念，它是研究抽样推断的基础。

7.1.1 总体和样本

1. 总体

总体(population)也称全及总体，是指所研究现象的整体，即包括所要调查的所有单位。例如，“做一做”中，从5 000名大学生中抽取学生进行调查，以推断学生的午餐花费。这5 000名学生是全及总体，一般用英文大写字母N表示总体的单位数，取$N=$5 000人。全及总体按其各单位性质的不同，可以分为变量总体和属性总体两类。对于一个总体来说，若被研究的单位标志属于品质标志，则该总体为属性总体，若被研究的单位标志属于数量标志，则该总体为变量总体。

2. 样本

样本(sample)又称子样，它是从全及总体中随机抽取出来，作为代表这一总体的那部分单位组成的集合体。样本的单位数是有限的，相对来说，它的数目比较小，一般用英文小写字母n表示样本的单位数，如$n=100$人。

作为推断对象的总体是确定的，而且是唯一的。但作为观察对象的样本就不是这样。从一个总体可以抽取很多个样本，每次可能抽到哪个样本不是确定的，也不是唯一的，而是可变的。

7.1.2 参数和统计量

1. 参数

根据总体各单位的标志值或标志属性计算的，并反映总体数量特征的综合指标称为全及指标。全及指标是总体变量的函数，其数值由总体各单位的标志值决定。由于总体是唯一确定的，因此，全及指标也是唯一确定的，所以也称参数(parameter)。常用的参数有总体平均数、总体成数、总体方差和总体标准差。

(1) 总体平均数。对于变量总体，由于各单位的标志用数量表示，因此可以计算总体平均数，通常用$\bar{X}$表示。

设X为总体的某一变量，其N项变量值为X_1，X_2，…，X_n，则总体平均数为

$$\bar{X}=\frac{X_1+X_2+\cdots X_n}{N}=\frac{\sum X}{N} \tag{7-1}$$

(2) 总体成数。对于属性总体，由于各单位标志不能用数量表示，因此总体参数常以成数或比例表示。通常以P表示总体中具有某种标志表现，即“是”的单位数在总体单位数中所占的比例；以Q表示不具有某种标志表现，即“非”的单位数所占的比例。

设总体N个单位中，有N_1个单位具有某种标志表现，N_0个单位不具有某种标志表现，且$N=N_1+N_0$，则总体成数为

$$P=\frac{N_1}{N} \tag{7-2}$$



$$Q=\frac{N_0}{N}=\frac{N-N_1}{N}=1-P \tag{7-3}$$

【实例 7.1】 亚明灯泡厂生产的 10 000 只灯泡中，有 9 550 只是合格品，有 450 只是不合格产品，则总体成数即灯泡的合格率和不合格率分别为

$$P=\frac{N_1}{N}\times100\%=\frac{9\ 500}{10\ 000}=95.5\%$$

$$Q=\frac{N_0}{N}\times100\%=\frac{450}{10\ 000}\times100\%=4.5\%$$

需要指出，统计上把只表现为“是”与“非”的标志称为是非标志，也称交替标志(如性别标志等)。

(3) 总体方差和总体标准差。就变量总体而言，总体方差和总体标准差计算公式分别为

$$\sigma^2=\frac{\sum(X-\bar{X})^2}{N} \tag{7-4}$$

$$\sigma=\sqrt{\frac{\sum(X-\bar{X})^2}{N}} \tag{7-5}$$

在属性总体条件下，则可以把“是”与“非”两种标志表现进行量化处理。用“1”表示“是”，即具有某种标志表现；用“0”表示“非”，即不具有某种标志表现。那么“是”的成数就可视为是非标志的平均数，从而计算出属性总体的方差和标准差，即

$$\sigma_P^2=(1-P)^2P+(0-P)^2Q=Q^2P+P^2Q=PQ(P+Q)=PQ=P(1-P) \tag{7-6}$$

$$\sigma=\sqrt{P(1-P)} \tag{7-7}$$

属性总体的方差和标准差的具体计算过程见表 7-2。

表 7-2 属性总体的方差和标准差的计算

变量 X	成数或频率 $\frac{f}{\sum f}$	变量与频率之积 $X\cdot\frac{f}{\sum f}$	离差平方 $(X-P)^2$	离差平方与频率之积 $(X-P)^2$
1(是)	P	P	$(1-P)^2$	$(1-P)^2P$
0(非)	Q	0	$(0-P)^2$	$(0-P)^2Q$
合计	1	P	—	$P\ (1-P)$

【实例 7.2】 仍以实例 7.1 的有关资料为例，计算其总体方差和标准差分别为

$$\sigma_p^2=P(1-P)\times100\%=95.5\%\times(1-95.5\%)\times100\%\approx4.3\%$$

$$\sigma_p=\sqrt{P(1-P)}\times100\%=\sqrt{95.5\%\times(1-95.5\%)}\times100\%\approx20.73\%$$

在抽样调查中，总体参数的意义和计算方法是明确的，但参数的具体数值事先是未知的，需要用抽样估计它。

2. 统计量

根据样本各单位标志值计算出的反映样本特征的指标称为统计量(statistic)，也称作样本指标，它是用来估计总体参数的。与总体参数相对应，统计量主要有样本平均数、样



本成数、样本方差和样本标准差等。

(1) 样本平均数。设 x 为样本的某一变量，其 n 项变量值为 x_1，x_2，…，x_n，则样本平均数为

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{\sum x}{n} \tag{7-8}$$

式中：$\bar{x}$ 为样本平均数。

(2) 样本成数。设样本的 n 个单位中有 n_1 个单位具有某种标志表现，n_0 个不具有某种标志表现，且 $n=n_1+n_0$，则样本成数为

$$p=\frac{n_1}{n} \tag{7-9}$$

$$q=\frac{n_0}{n}=\frac{n-n_1}{n}=1-p \tag{7-10}$$

式中：p、q 为样本成数。

【实例 7.3】 在"做一做"中，抽样调查了 100 名学生，其中，午餐花费在 4 元以下的学生有 8 人，在 4 元以上的人数为 92 人，则样本成数即午餐花费在 4 元以下人数比例和午餐花费在 4 元以上人数比例分别为

$$p=\frac{n_1}{n}\times 100\%=8\%$$

$$q=\frac{n_0}{n}=\frac{n-n_1}{n}=1-p=92\%$$

(3) 样本方差和样本标准差。它们是说明样本内各单位标志值变异程度的指标。

样本数量标志的方差 s^2 和标准差 s 的计算公式分别为

$$s^2=\frac{1}{n}\sum(x-\bar{x})^2 \tag{7-11}$$

$$s=\sqrt{\frac{\sum(x-\bar{x})^2}{n}} \tag{7-12}$$

样本是非标志的方差 s^2 和标准差 s 的计算公式分别为

$$s_P^2=P(1-P) \tag{7-13}$$

$$s_P=\sqrt{P(1-P)} \tag{7-14}$$

【实例 7.4】 以实例 7.3 资料为例，可计算出样本方差和样本标准差分别为

$$s_P^2=P(1-P)\times 100\%=8\%\times 92\%\times 100\%=7.36\%$$

$$s_P=\sqrt{P(1-P)}\times 100\%\approx 27.13\%$$

统计量作为样本变量的函数，其计算方法是确定的，但它的取值会随着样本的不同而变化，所以统计量本身也是随机变量。

7.1.3 样本容量和样本个数

样本容量和样本个数是两个互相联系但又完全不同的概念。样本容量是指一个样本所包含的单位数。一个样本应该包含多少单位最合适，是抽样设计必须认真考虑的问题。必须结合调查任务的要求，以及总体标志值的变异情况来考虑。样本容量的大小不但关系到



抽样调查的效果，而且关系到抽样方法的应用。通常将样本单位数不少于30个的样本称为大样本，单位数不及30个的样本称为小样本。社会经济统计的抽样调查多属于大样本调查。

样本个数又称样本可能数目，是指从一个总体上可能抽取的样本个数。一个总体可能抽取多少样本，和样本容量及抽样方法等因素都有关系，是一个比较复杂的问题。一个总体有多少样本，样本统计量就有多少种取值，从而形成统计量的分布。统计量的分布是抽样推断的基础。虽然在实践中只抽取个别或少数样本，但要判断所取样本的可能性就必须联系到全部可能取样本数目所形成的分布。

7.1.4 重复抽样和不重复抽样

重复抽样又叫重置(with replacement)抽样，是指从所研究的现象总体中，按照随机原则，抽取一个样本单位进行调查登记之后，把这个样本再放回总体中，然后从所研究的总体中抽取样本单位。这样，总体单位数不变，设它为N。在进行第二次和第三次抽取样本时，已经被抽中的样本，仍然有同等的机会再被抽中。

不重复抽样又叫不重置(without replacement)抽样，是指从所研究的现象总体中，按照随机原则，抽取一个样本单位进行调查登记之后，不再把这个样本放回去。这样，所研究现象的总体单位数将逐步减少，即总体单位数在不断起变化。

由于重复抽样和不重复抽样在抽样方法上不同，抽取的样本个数也不相同。从总体N个单位中，用重复抽样的方法，随机抽样取n个单位构成一个样本，则共可抽取N^n个样本。用不重复抽样的方法，抽取同样多个单位的样本，则全部可能抽取的样本数目为$N(N-1)(N-2)\cdots(N-n+1)$个。因此，在相同样本容量的条件下，重复抽样的样本个数总是大于不重复抽样的样本个数。

7.2 抽样误差

7.2.1 抽样误差的概念

抽样误差(sampling error)是指由于随机抽样的偶然因素，使各单位的结构不足以代表总体各单位的结构，而引起抽样指标和全及指标之间的绝对离差。例如，抽样平均数与总体平均数的绝对离差，抽样成数与总体成数之间的绝对离差等。例如，从5 000名学生中抽取100人调查午餐花费，其平均花费为6.08元，但其他4 900人中有的可能花费恰好是6.08元，但绝大部分的花费与样本的平均花费6.08元有差别。样本指标与总体指标之间存在绝对离差，这就是抽样误差。

抽样调查中误差有以下两个来源：一个是登记性误差，即在调查过程中，由于主客观原因而引起的误差；另一个是代表性误差，即样本各单位的结构情况不足以代表总体特征而引起的误差。引起代表性误差发生的两种情况：第一，非随机的代表性误差；第二，随机性误差。在第2章已经说过，在抽样调查中，由于被抽选的样本各种各样，被抽中的样本内部各单位标志的构成比例和总体特征有差异，就会出现或大或小的偶然性误差。这种

误差理论上或实际上都是不可避免的，只能做到尽量减少这种误差。抽样误差主要是指这种随机性误差，它可以精确计算，并加以控制。

影响抽样误差大小的因素主要有以下几个方面。

(1) 样本单位数。在其他条件不变的情况下，样本单位数越大，抽样误差越小；样本单位数越小，抽样误差就越大。这是因为，随着样本单位数的扩大，样本结构就越能反映总体的结构，样本指标就越能代表总体相应的数量特征。当样本扩大到等于全及总体时，则抽样调查就变为全面调查，代表性问题不再存在，因而也就没有抽样误差可言了。

(2) 总体内各单位被研究标志的变异程度。在其他条件不变的情况下，总体被研究标志的变异程度越大，抽样误差也越大；总体被研究标志的变异程度越小，抽样误差也越小。这是因为，总体标志变异程度小，表示总体各单位标志值之间的差异小，则样本指标与总体指标之间的差异也可能小。如果总体变异程度为零，即总体内各单位的标志值都相等，样本指标就完全等于总体指标，此时抽样误差也会为零。

(3) 抽样方法。重复抽样和不重复抽样是两种不同的抽样方法，抽样方法的不同，抽样误差的大小也不同。一般来说，不重复抽样的抽样误差要小于重复抽样的抽样误差。

(4) 抽样组织形式。采用不同的抽样组织形式，也会有不同的抽样误差，因为不同的抽样组织形式所抽出的样本对于总体的代表性不同。例如，通常情况下，按照机械抽样和类型抽样方式组织抽样调查，由于经过了排队和分类，可以缩小标志变异程度，因而抽取相同数目的调查单位，其抽样误差会比简单随机抽样方式的抽样误差小。

7.2.2 抽样平均误差

抽样平均误差是反映抽样误差一般水平的指标。通常用抽样平均数的标准差或抽样成数的标准差，作为衡量其抽样误差一般水平的尺度。按照标准差的一般意义，抽样平均数(或成数)的标准差是按抽样平均数(或成数)与其平均数的离差平方和计算的，但由于抽样平均数的平均数等于总体平均数，而抽样成数的平均数等于总体成数，则抽样指标的标准差恰好反映了抽样指标和总体指标的平均离差程度。相关计算公式为

$$\mu_{\bar{x}}=\sqrt{\frac{\sum(\bar{x}-\bar{X})}{M}} \tag{7-15}$$

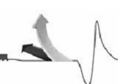
$$\mu_{p}=\sqrt{\frac{\sum(p-P)^2}{M}} \tag{7-16}$$

式中：$\mu_{\bar{x}}$ 为抽样平均数的平均误差；μ_p 为抽样成数的平均误差；M 为全部可能的样本数目。

这些公式表明了抽样平均误差的关系。但是由于总体平均数和总体成数并不知道，而且也无法计算全部样本的抽样指标值，所以按上述公式计算抽样平均误差在实际上是不可能的。在实用中可以通过其他方法加以推算。下面分别就抽样平均数和抽样成数的抽样平均误差的计算问题加以讨论。

1. 抽样平均数的平均误差

抽样平均数的平均误差分重复抽样和不重复抽样两种情况。



（1）重复抽样的抽样平均数的平均误差计算公式为

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}} \tag{7-17}$$

式中：$\mu_{\bar{x}}$ 为抽样平均数的平均误差；σ 为全及总体标准差(即均方差)；n 为抽样数目。

但是，σ 是全及总体标准差，这是不知道的。如果知道，就无须进行抽样调查了。在这种情况下可用样本的方差 s^2 代替，其计算公式为

$$s^2=\frac{\sum(x-\bar{x})^2}{n} \tag{7-18}$$

这是简单式方差的计算，如果是加权式，则计算公式为

$$s^2=\frac{\sum(x-\bar{x})^2 f}{\sum f} \tag{7-19}$$

（2）不重复抽样的抽样平均数的平均误差计算公式为

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)} \tag{7-20}$$

式中：N 为全及总体单位数。

当全及总体单位数 N 很大时，$N-1$ 接近于 N，即可用 N 代替，则式(7-20)可简化为

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} \tag{7-21}$$

由此可见，不重复抽样的抽样平均数的平均误差公式，与重复抽样的抽样平均数的平均误差公式的差别，只在于不重复抽样的抽样平均数的平均误差公式的根号下，乘以一个修正系数 $\left(1-\frac{n}{N}\right)$。因为 $\frac{n}{N}$ 总是一个正小数，用 1 减去 $\frac{n}{N}$，其值总是小于 1。而 $\frac{\sigma^2}{n}$ 乘上一个小于 1 的修正系数 $\left(1-\frac{n}{N}\right)$，其值总比 $\frac{\sigma^2}{n}$ 的值小，即 $\frac{\sigma^2}{n}>\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)$。

因此，不重复抽样所得到的抽样平均数的平均误差，比重复抽样所求得的抽样平均数的平均误差小。即

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}>\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} \tag{7-22}$$

【实例 7.5】 在“做一做”中，在张毅所在的分校区 5 000 名学生中，按简单随机抽样方式抽取 100 名学生调查，其平均午餐花费为 6.08 元，标准差为 1.97 元，根据重复抽样和不重复抽样的方法分别计算抽样平均误差。

（1）按重复抽样计算抽样平均误差：

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}\approx 0.197(\text{元})$$

（2）按不重复抽样计算抽样平均误差：

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}\approx 0.195(\text{元})$$

由上述计算结果可以看出，不重复抽样的抽样平均数的平均误差比重复抽样的抽样平

均数的平均误差小。在统计工作的实践中，一般采用不重复抽样的方法进行抽样调查。但其抽样平均误差的计算公式中带有一个修正系数$\left(1-\frac{n}{N}\right)$，计算起来较为麻烦，为了计算简便，一般采用重复抽样的抽样平均误差公式，这样计算并不影响抽样平均误差的准确性。

2. 抽样成数的抽样平均误差

（1）重复抽样的抽样成数的平均误差计算公式为

$$\mu_p=\sqrt{\frac{P(1-P)}{n}} \tag{7-23}$$

式中：μ_p 为抽样成数的抽样平均误差；P 为成数；n 为抽样单位数。

（2）不重复抽样的抽样平均误差计算公式。不重复抽样的抽样平均误差计算，是在重复抽样平均误差的公式中，乘上一个修正系数$\left(1-\frac{n}{N}\right)$，即

$$\mu_p=\sqrt{\frac{P(1-P)}{n}\left(1-\frac{n}{N}\right)} \tag{7-24}$$

式中：N 为全及总体单位数。

式(7-23)和式(7-24)中的 P 是总体成数，这是不知道的；而 $P(1-P)$是总体方差，它等于 s^2，所以 $P(1-P)=s^2$也是不知道的。在这种情况下，可用样本的成数 P 和样本方差 $P(1-P)$代替，或者用过去类似调查的成数和成数方差代替。成数方差的最大值为

$$P(1-P)=0.5\times(1-0.5)=0.5\times0.5=0.25$$

这就是说，总体内部两种情况各占 50%时，成数方差的最大值为 0.25。

【实例 7.6】 在“做一做”中，从 5 000 学生中随机抽取 100 名学生调查，结果有 8 人午餐花费在 4 元以下，计算午餐花费在 4 元以下人数比例的抽样平均误差。

（1）按重复抽样计算抽样平均误差：

$$\mu_p=\sqrt{\frac{P(1-P)}{n}}\times100\%\approx2.71\%$$

（2）按不重复抽样计算抽样平均误差：

$$\mu_p=\sqrt{\frac{P(1-P)}{n}\left(1-\frac{n}{N}\right)}\times100\%\approx2.68\%$$

以上计算结果表明，采用不重复抽样的方法算出的成数抽样平均误差，比采用重复抽样的方法算出的成数抽样平均误差小。

7.2.3 抽样极限误差

1. 抽样极限误差的概念

抽样极限误差又称抽样允许误差，是指样本指标与总体指标之间产生抽样误差被允许的最大可能范围，它是根据所研究对象的变异程度和分析任务的要求来确定的可允许的误差范围，凡是在这个范围内的数字都算有效，统计上把这种可允许的误差范围称为抽样极



限误差。根据样本指标估计总体指标必然会产生误差，但误差不能太大，否则会使抽样估计失去意义；但误差也不是愈小愈好，因为这样就要增加样本单位数，必然会提高很多费用，所以要确定一个误差范围。

抽样极限误差通常用样本指标可允许变动的上限或下限与总体指标的绝对离差表示。设 Δx 和 Δp 分别表示抽样平均数和抽样成数的抽样极限误差，则有

$$\Delta\bar{x}=|\bar{x}-\bar{X}| \tag{7-25}$$

$$\Delta p=|p-P| \tag{7-26}$$

公式(7-25)和公式(7-26)可变换为下列不等式：

$$\bar{x}-\Delta\bar{x}\leqslant\bar{X}\leqslant\bar{x}+\Delta\bar{x} \tag{7-27}$$

$$p-\Delta p\leqslant P\leqslant p+\Delta p \tag{7-28}$$

公式(7-27)和公式(7-28)表明总体平均数 $\bar{X}$ 以抽样平均数 $\bar{x}$ 为中心，在 $\bar{x}\pm\Delta\bar{x}$ 的范围内变动，总体成数 P 也是如此。这个变动范围称为估计区间。

2. 抽样误差的概率度

抽样极限误差与抽样平均误差不同，它是人为规定的，可大可小。而抽样平均误差是计算出来的，在一定的抽样方式中是唯一的。

基于概率估计的要求，抽样极限误差通常需要以抽样平均误差 $\mu_{\bar{x}}$ 或 μ_p 为标准单位来衡量。把极限误差 $\Delta\bar{x}$ 或 Δp 分别除以 $\mu_{\bar{x}}$ 或 μ_p，得出相对数 t，表示误差范围为抽样平均误差的 t 倍。t 是测量估计可靠程度的一个参数，称为抽样误差的概率度，用公式表示为

$$t=\frac{\Delta\bar{x}}{\mu_{\bar{x}}}=\frac{|\bar{x}-\bar{X}|}{\mu_{\bar{x}}} \tag{7-29}$$

$$t=\frac{\Delta p}{\mu_p}=\frac{|p-P|}{\mu_p} \tag{7-30}$$

抽样极限误差也可以表示为抽样平均误差的若干倍，其倍数即概率度 t。用公式表示为

$$\Delta\bar{x}=|\bar{x}-\bar{X}|=t\cdot\mu_{\bar{x}} \tag{7-31}$$

$$\Delta p=|p-P|=t\cdot\mu_p \tag{7-32}$$

上述关系式表明，抽样极限误差 Δ 可以用 t 倍的抽样平均误差 m 表示。在抽样平均误差为一定的条件下，概率度 t 的数值越大，则极限误差 Δ 越大，用样本指标估计总体指标的可靠程度也就越高，估计的精确程度就越低；反之若 t 的数值越小，则 Δ 越小，抽样估计的可靠程度也就越低，估计的精确程度就越高。

7.3 抽样推断的方法

7.3.1 参数估计

抽样估计是指利用实际调查计算的样本指标值来估计相应的总体指标的数值。由于总体指标是表明总体数量特征的参数，所以也称为参数估计。总体参数估计有点估计(point

estimation)和区间估计(interval estimation)两种。

1. 点估计

点估计的基本特点是，根据总体指标的结构形式设计样本指标作为总体参数的估计量，并以样本指标的实际值作为相应总体参数的估计值(例如，以样本平均数的实际值作为相应总体平均数的估计值，以样本成数的实际值作为相应总体成数的估计值等)。我们所以做这样的考虑，是基于我们对所研究的总体指标的具体指标值虽然不知道，但是对于它的指标结构形式都是清楚的。

例如：抽选 6 000 名大学生，对其实际月消费支出调查，结果表明这 6 000 名大学生的月平均消费支出为 826 元，就可以推断说，全体大学生的月消费支出为 826 元。

对总体参数做估计的时候，总是希望估计是合理或优良的。那么优良估计的标准是什么呢？所谓优良估计，总是从总体上评价的，其标准有 3 个方面。

(1) 无偏性(unbiasedness)。即以抽样指标估计总体指标要求抽样指标值的平均数，等于被估计的总体指标值本身。就是说，虽然每一次的抽样指标值和总体指标值之间都可能有误差，但在多次反复的估计中，各个抽样指标值的平均数应该等于所估计的总体指标值本身，即对抽样指标的估计，平均说来是没有偏误的。

从 7.2 节已经知道，抽样平均数的平均数等于总体平均数，抽样成数的平均数等于总体成数，即

$$E(\bar{x}) = \bar{X} \tag{7-33}$$

$$E(p) = P \tag{7-34}$$

这说明以抽样平均数作为总体平均数的估计量，以抽样成数作为总体成数的估计量，是符合无偏性原则的。

(2) 一致性(consistency)。当以抽样指标估计总体指标要求当样本的单位数充分大时，抽样指标也充分地靠近总体指标。就是说，随着样本单位数 n 的无限增加，抽样指标和未知的总体指标之差的绝对值小于任意小的数，它的可能性也趋近于必然性，即实际上是几乎肯定的。我们知道，抽样平均数和抽样成数的抽样平均误差和样本单位数的平方根成反比例变化，样本单位数愈多则平均误差便愈小，当样本单位数接近于总体单位数时，平均误差也就接近于零。也就是说，抽样平均数和抽样成数作为总体平均数和总体成数的估计量，是符合一致性原则的。

(3) 有效性(efficiency)。以抽样指标估计总体指标要求作为优良估计量的方差，应该比其他估计量的方差小。例如，用抽样平均数或总体某一变量值来估计总体平均数，虽然两者都是无偏的，而且在每一次估计中，两种估计量和总体平均数都可能有离差，但样本平均数更靠近于总体平均数的周围，平均说来其离差比较小。所以对比说来，抽样平均数是更有效的估计量。

总体参数点估计方法的优点是简便、易行、原理直观，常为实际工作者所采用。但也有不足之处，这种估计没有表明抽样估计的误差，更没有指出误差在一定范围内的概率保证程度有多大，要解决这个问题，就必须采用总体参数的区间估计方法。



2. 区间估计

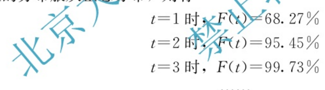
1) 区间估计的概念和要素

区间估计是在一定的概率把握程度下，根据样本指标和抽样极限误差估计总体指标所在可能范围的方法。这个范围通常用一个最低限和一个最高限构成的区间表示，并以一定的概率把握程度保证总体指标的估计值在这两个数值构成的区间之内。在总体指标的区间估计公式中，有两个要素，一个是置信区间，另一个是置信概率。

(1) 置信区间(confidence interval)又称为估计区间，是指由低限到高限两个数值所构成的可能范围，由样本指标和极限误差组成。置信区间是一个随机区间，这是由其组成要素的随机性决定的。置信区间越小，允许的抽样误差范围也越小，则估计的准确性越高；反之，估计的准确性越低。

(2) 置信概率(confidence probability)是指区间估计的概率保证程度，也称为置信度。它是表明样本指标与总体指标的误差不超过一定范围的概率有多大，在样本指标的基础上，运用极限误差进行估计，总是要和一定的概率保证程度联系在一起。因为我们采用的是随机抽样，抽样误差是一个随机变量，则不能保证误差不超过一定范围是一个必然事件，而只能给予一定程度的概率保证。所以，在进行抽样估计时，不仅要考虑极限误差的大小，而且要考虑被估计的数值落在这个范围的概率的大小。前者是估计的准确性问题，后者是估计的可靠性问题，两者紧密联系且不可分割。

数理统计已证明，概率保证程度与概率度 t 密切联系，它是 t 的函数，表示为 $F(t)$，抽样平均数的分布服从正态分布，则有



$t=1$ 时，$F(t)=68.27\%$

$t=2$ 时，$F(t)=95.45\%$

$t=3$ 时，$F(t)=99.73\%$

……

从“正态分布概率表”(见附录 A)中，可以看出，$F(t)$随着 t 的增大而增大，即 t 增大了，可靠性也增大了。但抽样极限误差也扩大了，估计的准确性也降低了。

2) 总体平均数和总体成数的区间估计。

(1) 总体平均数的估计。总体平均数的估计就是用抽样平均数来估计总体平均数。

【实例 7.7】 在“做一做”中，对抽选的 100 名学生进行调查，资料见表 7 - 1。

根据这些资料，要求按重复抽样方法以 95%的概率估计张毅所在分校区学生平均午餐花费的范围。

解：

由已知条件计算抽样平均数、标准差：

$$\bar{x}=\frac{\sum xf}{\sum f}\approx 6.08(\text{元})$$

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}\approx 1.97(\text{元})$$

抽样平均误差为

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}\approx 0.197(\text{元})$$

抽样极限误差为

$$\Delta x=1.96\times 0.197\approx 0.39(\text{元})$$

所以，在95%的概率保证程度下，张毅所在分校区学生平均午餐花费的范围在5.69～6.47元。

$$\text{下限}=\bar{x}-\Delta\bar{x}=5.69(\text{元})$$

$$\text{上限}=\bar{x}+\Delta\bar{x}=6.47(\text{元})$$

(2) 总体成数的估计。总体成数的估计就是用抽样成数来估计总体成数。

【实例7.8】 利用“做一做”的资料，同样以95%的概率估计分校区午餐花费在4元以下的学生人数所占的比例的范围。

解：

根据资料可得

样本午餐花费在4元以下的学生人数所占的比例为8%

抽样平均误差 $\mu_p=\sqrt{\frac{P(1-P)}{n}}\times 100\%\approx 2.71\%$

抽样极限误差 $\Delta p=1.96\times 2.71\%\approx 5.31\%$

$$\text{下限}=p-\Delta p=2.69\%$$

$$\text{上限}=p+\Delta p=13.31\%$$

所以，在95%的概率保证程度下，分校区午餐花费在4元以下的学生人数所占的比例的范围在2.69%～13.31%。由于分校区有学生5 000人，所以以95%可靠程度推断，分校区午餐花费在4元以下的学生人数有135～666人。

7.3.2 样本容量的确定

1. 确定必要样本容量的意义

样本单位数是决定抽样误差大小的直接因素，因此，在组织抽样调查时，必须事先确定样本单位数。样本单位数越多，样本的代表性越大，抽样误差越小，抽样估计就越可靠。但样本容量过多会增加不必要的人力、物力和费用开支，造成浪费。样本容量减少，会使抽样误差增大，达不到所需要的准确程度。可见，合理确定必要的样本单位数，以保证取得较好的抽样估计效果，是组织抽样调查的重要问题。

2. 影响样本容量的因素

(1) 总体各单位标志变异程度。总体标志变异程度大，要求样本容量大；反之，总体标志变异程度小，样本容量小。

(2) 抽样方法。在其他条件相同的情况下，重复抽样要比不重复抽样多抽取一些样本单位。



（3）抽样组织形式。一般来说，类型抽样和等距抽样的样本容量要小于简单随机抽样的样本容量。

（4）极限误差的大小。如果允许误差大，样本容量就小；如果允许误差小，样本容量就大。

（5）抽样估计的可靠程度即概率 $F(t)$的大小。如果估计的可靠程度要求越高即 $F(t)$越大，样本容量就越多；如果估计的可靠程度要求越低即 $F(t)$越小，样本容量就越少。

以上影响样本容量的因素，可以从样本容量的计算公式加以印证。

3. 必要样本容量的计算公式

（1）平均数的必要样本容量的计算公式。

在重复抽样的条件下为

$$n=\frac{t^2\sigma^2}{\Delta\overline{x}^2} \tag{7-35}$$

在不重复抽样的条件下为

$$n=\frac{Nt^2\sigma^2}{N\Delta^2\overline{x}+t^2\sigma^2} \tag{7-36}$$

（2）成数的必要样本容量的计算公式。

在重复抽样的条件下为

$$n=\frac{t^2P(1-P)}{\Delta^2 p} \tag{7-37}$$

在不重复抽样的条件下为

$$n=\frac{Nt^2P(1-P)}{N\Delta^2 p+t^2P(1-P)} \tag{7-38}$$

【实例 7.9】 对某罐头厂生产的罐头质量进行抽样调查，抽样极限误差为 5%，概率为 0.954 5，并知过去进行同样抽样调查，其不合格率为 10%，试求必要的样本容量。

根据题意可知：$t=2$，$P=90\%$，$p=5\%$。

在重复抽样的条件下：

$$n=\frac{t^2P(1-P)}{\Delta^2 p}=\frac{2^2\times 90\%\times(1-90\%)}{(5\%)^2}=144(\text{盒})$$

4. 计算必要样本容量应注意的问题

（1）上述公式计算的样本容量是最低的，也是最必要的样本容量。

（2）用上述公式计算样本容量时，一般情况下，总体方差是未知的，处理方法与“抽样平均误差计算公式”相同。

（3）如果进行一次抽样调查，同时对总体平均数和成数进行估计，此时就会计算出两个样本容量，并且二者通常情况下又是不相等的。那么，为了同时满足两个估计的要求，应该在两个样本容量中选择较大的一个进行抽样。

（4）上述公式计算的样本容量不一定是整数。如果带有小数，不允许采取四舍五入的办法化为整数，而是用比这个小数稍大的邻近整数代替。

7.3.3 抽样的组织形式

抽样推断是根据事先规定的要求而设计的抽样调查组织，并以所获得的这一部分实际资料为基础，进行推理演算做出结论。如何科学地组织抽样调查，保证随机抽样条件的实现，并合理有效地取得各项实际数据，成为抽样中一个至关重要的问题。本节主要介绍几种常用的抽样组织形式，简单随机抽样(simple random sample)、分层抽样(stratified sampling)、等距抽样(systematic sampling)、整群抽样(cluster sampling)和多阶段抽样(multi-stage sampling)等。

1. 简单随机抽样

简单随机抽样又称纯随机抽样。它是按随机原则直接从总体 N 个单位中抽取 n 个单位作为样本。不论是重复抽样或不重复抽样，都要保证每个单位在抽选中都有相等的中选机会。简单随机抽样是最基本、最简单的抽样组织形式，它适用于均匀总体，即具有某种特征的单位均匀地分布于总体的各个部分，使总体的各部分都是同分步的。通常，抽取样本可采用下面的直接抽取法、抽签法和随机数表法。

(1) 直接抽取法是从调查总体中直接随机抽取样本进行调查。这种方法适合对集中于某个较小空间的总体进行抽样，如对存放于仓库的同类产品直接随机抽出若干产品为样本进行质量检查。

(2) 抽签法是将总体各单位编上序号并将号码写在外形相同纸片上掺和均匀后，再从中随机抽取。被抽中的号码所代表的单位就是随机样本，直到抽够预先规定的样本数目为止。

(3) 随机数表法是先把总体各单位编号，根据编号的最大数(即总体单位数位数确定使用随机数表中若干列或若干行数字)，从任意行或任意列的第一个数字起，可以向任何方向去数，遇到属于总体单位编号范围内的号码就定为样本单位，直到抽取足够的预定的样本单位数为止。随机数表(见附录C)是由0～9组成的表，这10个数字排列完全是随机的。较大的随机数表有美国兰德公司1955年编制出版的100万数字表和肯德尔(Kendall)与史密斯(Smith)在1938年编制出版的10万数字表。随机数表法比抽签简化，它免去了对每个总体单位做签的过程，这在总体单位数 N 较大时，尤其显示出它的便捷。

简单随机抽样是随机抽样最基本的方法，也是其他随机抽样法的基础。它完全符合随机原则。且应用简单易行，但这种方法的不足在于：在总体很大的情况下使用时，编号工作量繁重；当总体单位差异程度较大时，必须使用样本容量充分大才能保证样本推断总体的可靠程度和准确程度；所抽取的样本在总体中的分布或是过于集中，或是过于分散，很不均匀，给实际调查带来困难，在实践应用中有不便之处。

2. 分层抽样

分层抽样又称类型抽样。它是先对总体各单位按主要标志加以分组，然后从各组中按随机原则抽选一定单位构成样本。

例如，调查城市居民收入情况，为了保证样本单位能容纳各种类型的居民，可将城市



居民按其工作单位的所有制性质分组，划分为全民所有制、集体所有制、个体所有制、联营制等类型，然后在每个类型中随机抽选若干样本单位进行调查，或者按居民的职业分组来抽选样本单位。这样就能够保证样本有较充分的代表性。

将总体分成若干组后，在各组中抽选样本单位有以下 3 种方法。

第一，按各组的标志变异程度确定各组的应抽单位数。对于标志变异大的组，宜多抽一些单位进行调查；对标志变异小的组，宜少抽一些单位进行调查。这种分配方法称为样本单位最佳分配。

第二，按调查者主观意志任意确定各组应抽选单位数，这种分配方法称为随意分配。

第三，按统一的比例确定各组应抽选的单位数，这种方法称为比例分配。本节主要讨论按比例分配样本单位的方法。比例分配使下列条件存在：

$$\frac{n_1}{N_1}=\frac{n_2}{N_2}=\cdots=\frac{n_k}{N_k}=\frac{n}{N} \tag{7-39}$$

式中：k 为总体分成的组数；n_1，n_2，…，n_k为各组应抽选的单位数；N_1，N_2，…，N_k为各组中的总体单位数。

并且 $\sum n_i=n$，$\sum N_i=N$；因而使 $n_i=N_i\cdot\frac{n}{N}=n\cdot\frac{N_i}{N}$。

如果按比例分配样本单位，可以得到如下类型抽样误差公式。

（1）在重复抽样条件下：

$$\text{变量总体 } \mu_{\bar{x}}=\sqrt{\frac{\overline{\sigma_i^2}}{n}} \tag{7-40}$$

式中：$\overline{\sigma_i^2}=\frac{\sum\sigma_i^2 n_i}{n}$是各组方差的加权平均数，习惯上称为平均组内方差。

$$\text{属性总体 } \mu_p=\sqrt{\frac{\overline{P_i(1-P_i)}}{n}} \tag{7-41}$$

式中：$\overline{P_i(1-P_i)}=\frac{\sum P_i(1-P_i)n_i}{n}$是各组比率的方差加权平均数，是属性总体的平均组内方差。

（2）在不重复抽样条件下：

$$\text{变量总体 } \mu_{\bar{x}}=\sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1-\frac{n}{N}\right)} \tag{7-42}$$

$$\text{属性总体 } \mu_p=\sqrt{\frac{\overline{P_i(1-P_i)}}{n}\left(1-\frac{n}{N}\right)} \tag{7-43}$$

式中：总体参数 s_i^2 和 P_i是未知的，习惯用样本值 S_i^2 和 P_i代替。

【实例 7.10】 对某乡全部 4 000 亩土地按类型抽样了解该乡平均亩产和全乡小麦总产量，其中有平原地 3 000 亩，丘陵地区 1 000 亩，按 3‰ 比例抽取样本单位，调查结果见表 7－3。要求在 95％的概率保证下，对全乡小麦平均产量和总产量进行估计。

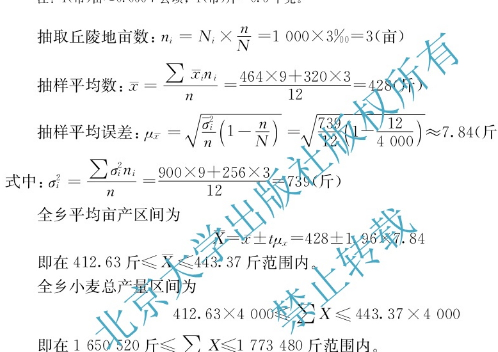
解：

抽取平原地亩数：$n_i=N_i\times\frac{n}{N}=3\ 000\times 3‰=9$（亩）

表 7-3　某村小麦产量抽样资料

土地类型	全部面积 (N_1)/亩	抽样面积 (n_i)/亩	抽样单位产量 (x_i)/(斤/亩)	各组抽样平均数 $(\overline{x_i})$/(斤/亩)	样本方差 s_i^2/斤
平原	3 000	9	420、420、450 460、465、470 480、490、520	464	900
丘陵	1 000	3	300、320、340	320	256
合计	4 000	12	5 135	—	—

注：1(市)亩≈0.066 7 公顷；1(市)斤＝0.5 千克。



抽取丘陵地亩数：$n_i = N_i \times \dfrac{n}{N} = 1\ 000 \times 3‰ = 3$(亩)

抽样平均数：$\bar{x} = \dfrac{\sum \bar{x}_i n_i}{n} = \dfrac{464\times 9 + 320\times 3}{12} = 428$(斤)

抽样平均误差：$\mu_{\bar{x}} = \sqrt{\dfrac{\overline{\sigma_i^2}}{n}\left(1-\dfrac{n}{N}\right)} = \sqrt{\dfrac{739}{12}\left(1-\dfrac{12}{4\ 000}\right)} \approx 7.84$(斤)

式中：$\overline{\sigma_i^2} = \dfrac{\sum \sigma_i^2 n_i}{n} = \dfrac{900\times 9 + 256\times 3}{12} = 739$(斤)

全乡平均亩产区间为

$$\bar{X} = \bar{x} \pm t\mu_{\bar{x}} = 428 \pm 1.96 \times 7.84$$

即在 412.63 斤$\leqslant \bar{X} \leqslant$443.37 斤范围内。

全乡小麦总产量区间为

$$412.63 \times 4\ 000 \leqslant \sum X \leqslant 443.37 \times 4\ 000$$

即在 1 650 520 斤$\leqslant \sum X \leqslant$1 773 480 斤范围内。

另外，在计算必要抽样单位数目中，用 $\overline{\sigma}^2$ 取代简单随机抽样中的 σ^2，即可得到类型抽样 n 的数目。例如，在重复条件下的变量总体，它的样本单位应为

$$n = \frac{t^2\ \overline{\sigma_i^2}}{\Delta^2 \bar{x}} \tag{7-44}$$

有关区间估计的方法和步骤，与简单随机抽样一致。

3. 等距抽样

等距抽样也称机械抽样或系统抽样。它先按某一标志对总体各单位进行排队，然后依一定顺序和间隔抽取样本单位的一种抽样组织。由于这种抽样是在各单位大小顺序排队基础上，再按某种规则依一定间隔取样，这样可以保证所取得的样本单位比较均匀地分布在总体的各个部分，有较高的代表性。

与类型抽样一样，在抽样前选择了某一标志划分总体，这是它们的一致之处。不一致之处如下：

第一，类型抽样选择标志是与研究目的有重要关系的标志，而机械抽样既能选择与研究目的有某种关系的标志，也能选择与研究目的无关的标志。例如，调查工人工资水平，



如用类型抽样，就必须选择与工资水平这一调查内容有密切联系的“工龄”标志将总体分组，分成“10 年以下”、“10～20 年”、“20 年以上”等组，如用机械抽样，可选择与调查内容无关的“姓氏笔画”或“身高”这些标志将所有工人按照顺序排列编号。

第二，类型抽样选择标志目的是将总体单位划分为有限个组(组数较少)，在每组中抽取 n_i 个样本单位。而机械抽样选择标志的目的是将总体单位顺序排列，划成 n 个相同单位部分或组(组数较多)，在每组中仅抽取一个样本单位。

第三，类型抽样选择分组标志可以是品质标志和数量标志。如调查居民收入，可按照“职业”品质标志分组，也可按照“工龄”数量标志分组。但是机械抽样只选数量标志排队。如上述选“姓氏笔画”、“身高”等数量标志将工人顺序排列编号。

机械抽样根据总体单位数(N)和需抽选的样本单位数(n)，可将总体单位划成 n 个相等部分，每部分都包括 $K=\dfrac{N}{n}$ 个单位，然后在第一个部分中抽选一个单位，依 K 个间隔依次在各部分都抽取一个单位组成样本。

机械抽样的重要工作是选取第一个样本单位。一般说来，确定的方法主要有以下 3 种。

第一种，随机取样。从第一部分包括的 K 个单位中随机确定任意 i 号单位为第一个样本单位，然后依次在第二部分中取($i+K$)号为第二个样本单位，第三部分取($i+2K$)号……第 n 部分取 $\{i+(n-1)K\}$ 号，直至取 n 个单位为止。由上可见，当第一个样本单位 i 号被确定后，其他各样本单位随之也被确定下来。

【实例 7.11】 某街道有 10 000 名职工，现需统计他们的收入情况，拟抽取 100 名职工进行调查。按照与研究目的有关的标志“工龄”，将 10 000 名职工顺序排列，将每名职工按 00001、00002……10000 编上相应号码，然后将 10 000 名职工分成 100 个相等部分，每部分 $K=\dfrac{N}{n}=\dfrac{10\ 000}{100}=100$ 名职工，第一部分包括 00001～00100 号，第二部分为 00101～00200 号……按随机原则在第一部分抽取 00014 号为第一个样本单位，那么根据依次加 $K=100$ 个间隔抽取样本单位的要求，其号码是 00114、00214、00314……09814、09914，即 100 个单位组成样本。

第二种，中位取样。中位取样与随机取样不同之处是第一个样本单位不是随机抽选，而是在第一部分的中间位置(或靠近中间位置)的号码为第一个样本单位号码，即$\dfrac{K+1}{2}$(或$\dfrac{K}{2}$)号码为第一个中选单位。依上例用中位取样，依次抽取的号码是 00050、00150、00250……09950。在城市住户调查中，按照有关标志排列进行中位取样的方法比随机取样方法代表性好，因此得到比较广泛的应用。

第三种，对称取样。对称取样是在中位取样的基础上改进的。它是在每部分的两侧按对称位置抽选样本单位。如果随机地在第一部分抽取第一个样本单位(i)号，那么在第二部分抽取第二个样本单位为($2K-i$)号，第三个样本单位则在($2K+i$)号，以后依次为($4K-i$)、($4K+i$)、($6K-i$)、($6K+i$)……即在单数部分抽取 K 的偶倍数加 i 号，在双数部分抽取 K 的偶倍数减 i 号。这样所抽选的结果使各部分的样本单位标志值大小具有相

互抵消而趋于平均数的趋向，使样本代表性增加，抽样误差减小。

仍以实例，如采用对称取样，从第一部分取第 00030 号($i=30$)为第一个样本单位，则在第二部分取第 00170 号($2K-i=200-30$)为第二个样本单位，依次抽取的号码是 00230、00370……09830、09970 等号，这种对称样本在按有关标志排列下能够取得比类型抽样更小的误差。因此，国家统计局要求农产量调查、城乡住户调查尽可能使用此方法。

机械抽样的抽样误差，直接计算很困难，可用下列方法处理。一是从理论上说，按无关标志排列的机械抽样，用简单随机抽样误差公式计算；按有关标志排列的机械抽样，用类型抽样误差公式计算。但在实践上，都用简单随机抽样误差公式计算。二是机械抽样都采用不重复方法抽选样本单位，因而都用不重复抽样误差公式计算。

机械抽样在总体单位标志数值差异程度比较大时运用，其优越性更能得到充分发挥。

4. 整群抽样

整群抽样又称集团抽样。它是将总体单位划分成 R 群(组)，每群(组)包括 M_i 个单位，然后从 R 群中随机地抽选若干群(r 群)，对每个抽中群的 M_i 个单位进行全面调查的一种组织方式。

例如，进行城市住户调查，将全市住户按居委会为基本单位分成 R 群，然后将各居委会按某个标志排队编号，或者等距或者随机地抽选 r 个居委会，对抽中的居委会的全部住户进行调查。

整群抽样的优点在于组织工作简单、搜集资料方便容易、调查费用较少，但是抽样误差一般都比较大。即使要得到同简单随机抽样相同的精确度，整群抽样也要调查大得多的样本单位。

整群抽样是以群为抽取基本单位，抽一次得到的包括 M_i 个单位为一群。每群大小可以相同(如批量生产的产品质量调查，每批数目大致相同)，也可以不同(如上例每个居委会住户数目不同)。通常我们调查的各群 M_i 相同时，用简单平均法计算样本指标；各群 M_i 不相同时，用加权平均法计算样本指标。

整群抽样都采用不重复抽样的方法，因此计算抽样误差公式都用不重复抽样公式。

$$\text{变量总体 } \mu_{\bar{x}}=\sqrt{\frac{\delta^2}{r}\left(\frac{R-r}{R-1}\right)} \tag{7-45}$$

式中：σ^2 为组间方差。

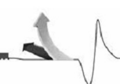
当每群大小相同时用简单式：

$$\delta^2=\frac{\sum(\overline{x_i}-\bar{x})^2}{r} \tag{7-46}$$

式中：$\overline{x_i}$ 为各群平均数，$\overline{x_i}=\dfrac{\sum x_i}{M_i}$；$\bar{x}$ 为样本平均数，$\bar{x}=\dfrac{\sum \overline{x_i}}{r}$。

当每群大小不同时用加权式(公式从略)。

$$\text{属性总体 } \mu_p=\sqrt{\frac{\delta^2}{r}\left(\frac{R-r}{R-1}\right)} \tag{7-47}$$



当每群大小相同时用简单式：

$$\delta^2 = \frac{\sum (p_i - p)^2}{r} \tag{7-48}$$

式中：p_i为各群抽样成数，$p_i = \frac{M_{i1}}{M_i}$；P 为样本成数，$p = \frac{\sum p}{r}$。

当每群大小不同时用加权式(公式从略)。

必要抽样群数 r 的确定公式类同简单随机抽样中有关公式。只不过现在确定的是 r，并用组内方差 σ^2 取代 σ^2。

$$r = \frac{Rt^2\delta^2}{R\Delta^2 + t^2\delta^2} \tag{7-49}$$

【实例 7.12】 某产品合计生产 3 000 批，每批产量大致相同，抽样检查产品合格率。要求其允许误差不超过 2%，若根据过去该产品质量资料，其各批群间方差为 3%。问在 95.45%($t=2$)可靠程度下需抽多少批产品为宜?

$$r = \frac{Rt^2\delta^2}{R\Delta^2 + t^2\delta^2} = \frac{3\,000 \times 2^2 \times 0.03}{3\,000 \times 0.02^2 + 2^2 \times 0.03} \approx 273(\text{批})$$

即要在全部产品中抽 273 批产品检查。

5. 多阶段抽样

多阶段抽样是在组织抽样时，不是从总体中一次直接抽取样本单位，而是把抽样过程分成几个过渡阶段进行，到最后才具体抽取样本单位。它是先抽大单位，再在大单位中抽小单位，在小单位中抽更小的单位，直到抽出最终能取得推断总体的基本单位为止。例如，进行全省企业职工家庭状况调查，先在全省中抽市，在抽中的城市的各部中抽选企业，最后在抽中的企业中抽选职工，进行职工家庭情况调查。

一般在总体很大、单位分布面积广，如从总体中采取一次直接抽取足够多的样本很困难时，可采用多阶段抽样。

多阶段抽样的具体步骤：首先，将调查总体各单位按一定标志分成若干集体，作为抽样的第一级单位，然后将第一级单位再分成若干小集体，作为抽样的第二级单位，依次类推，按研究问题的需要和现象本身的特点，分出第三级单位、第四级单位等；其次，依照随机原则，先在第一级单位中抽出若干单位作为第一级单位样本，然后在第一级被抽中的单位中再抽出第二级单位样本，依次类推，还可抽出第三级单位样本、第四级单位样本，由此形成了两阶段随机抽样、三阶段随机抽样或四阶段随机抽样等。

在多阶段抽样中，前几阶段的抽样，都类似整群抽样。每一阶段抽样都会存在抽样误差。为提高抽样指标的代表性，各阶段抽取群数的安排和抽样方式，都应注意样本单位的均匀分布。

首先，适当多抽第一阶段的群数，使样本单位在总体中得到均匀分布。但是，需要注意的是，样本过于分散则需要更多的人力和经费。

其次，根据方差的大小，考虑各阶段抽取群数的多少。对于群间方差大的阶段，应当适当多抽一些群；反之，则可少抽一些群。

最后，各阶段抽样时，可以根据条件将各种抽样组织方式灵活运用，而且尽可能利用现成资料。

以两阶段抽样为例，首先将总体划分为 R 组，而每组包含 M_i 个单位。第一步，从 R 组中随机抽取 r 组；第二步，再从中选的 r 组中分别随机抽取 m_i 个单位，构成一个样本。这种抽样就是两阶段抽样。其中，总体单位数 $N=M_1+M_2+\cdots+M_R$，各组的单位数 M_i 可以是相等的，也可是不等的。样本单位数 $n=m_1+m_2+\cdots+m_r$，各组抽取的样本单位可以是相等的，也可以是不等的。为简化起见，假定 R 组中各组的单位数相等，都为 M，则有 $N=RM$，而且从各组抽取的单位数也相等，都为 m，则有 $n=rm$。

两阶段抽样和类型抽样、整群抽样同样都须先对总体加以分组，再抽取单位，但它们之间却有明显的差别。类型抽样是从全部的分组中每组各抽取单位，它和两阶段抽样的区别在于第一阶段抽取了全部的组，而两阶段抽样在第一阶段只是随机地抽取部分的组。整群抽样是从全部的分组中随机抽取部分的组，然后对中选的全部单位进行调查，它和两阶段抽样的区别在于第二阶段抽取了中选组的全部的单位，而两阶段抽样在第二阶段只是在中选组中随机地抽取部分单位。所以，两阶段抽样在组织技术上是整群抽样和类型抽样的综合。

7.4 参数假设检验

7.4.1 假设检验的基本概念

假设检验(hypothesis testing)是利用样本的实际资料，检验事先对总体某些数量特征所做的假设是否可信的一种统计分析方法。它和参数估计一样，都是利用样本资料对总体特征进行某种推断。但二者推断角度不同，参数估计是根据样本指标以一定的把握程度估计总体参数取值范围，更准确地说，是对总体参数进行了区间估计；而假设检验则先对总体参数值提出一个假设，然后利用样本信息，以一定的概率水平判断这个假设是否成立。下面的一些例子，都是假设检验问题。

【实例 7.13】 某食品公司生产一种罐头，按标准每罐净重为 227 克，根据以往生产经验，罐头重量的标准差为 3 克。现随机抽查该公司产品 100 罐，测得平均净重为 228 克，判断这批罐头是否符合标准?

【实例 7.14】 某轮胎制造商在广告中声称，该公司生产的汽车轮胎在正常行驶条件下平均寿命高于 28 000 千米。检测部门随机挑选了 40 个轮胎进行测试，结果显示平均行驶里程 28 200 千米，样本标准差为 1 000 千米，问该制造商广告是否属实?

【实例 7.15】 某批发商欲从厂家购进一批灯泡，根据合同规定，灯泡的平均使用寿命不能低于 1 000 小时。从产品中随机抽取 100 只灯泡，测得平均寿命为 960 小时，标准差为 200 小时，请问批发商是否应该购买这批灯泡?

从直观上看，上面例子中样本指标数值与总体指标数值的要求有一些差异，但这种差异可能是由于抽样的随机性带来的，事实上也许并没有显著差异。究竟是否存在显著差异，可以先对总体设立一个假设，然后检验这个假设是否成立，这就是一个假设检验问



题。由此可见，假设检验是对我们所关心的，却又是未知的总体参数先做出假设，然后抽取样本，利用样本提供的信息对假设的正确性进行判断的过程。它是进行经济管理和决策的重要工具。

7.4.2 假设检验的步骤

一个完整的假设检验过程，通常包括以下 4 个步骤。第一，提出原假设和备择假设；第二，选取和计算检验统计量；第三，根据显著性水平查临界值；第四，进行比较并做出决策。

下面，我们分别对每一个步骤的内容加以分析和说明。

1. 提出原假设和备择假设

在统计学中，把需要通过样本推断其正确与否的命题称为原假设，用 H_0 表示。而与原假设相对立的假设便是备择假设，用 H_1 表示。

在对总体均值进行假设检验时，用 $\bar{x}_0$ 表示某特定的数值，所提出的假设共有 3 种情况。

第一种：$\begin{cases} H_0：\bar{X}=\bar{x}_0 \\ H_1：\bar{X}\neq\bar{x}_0 \end{cases}$ 第二种：$\begin{cases} H_0：\bar{X}\geqslant\bar{x}_0 \\ H_1：\bar{X}<\bar{x}_0 \end{cases}$ 第三种：$\begin{cases} H_0：\bar{X}\leqslant\bar{x}_0 \\ H_1：\bar{X}>\bar{x}_0 \end{cases}$

请注意，等号总是在原假设上。

第一种情况的假设检验，我们称为双侧检验(two-tailed test)。第二、第三种情况的假设检验，我们称为单侧检验(one-tailed test)。其中第二种称为左单侧检验，第三种称为右单侧检验。

当我们面临实际问题进行检验时，如何提出假设？可以按照下面 3 步来做。

首先，将所要检验的问题用符号表述。在实例 7.13 中要检验的问题是“罐头是否符合标准”，表述为 $\bar{X}=\bar{x}_0$；实例 7.14 中要检验广告是否属实，即汽车轮胎平均寿命是否高于 28 000 千米，表述为 $\bar{X}>\bar{x}_0$；实例 7.15 所关心的是批发商是否应该购买这批灯泡，而据合同，实际上检验的是灯泡的平均使用寿命是否不低于 1 000 小时，表述为 $\bar{X}\geqslant\bar{x}_0$。

其次，写出刚才所表述式子的余集。实例 7.13 的余集为 $\bar{X}\neq\bar{x}_0$，实例 7.14 的余集为 $\bar{X}\leqslant\bar{x}_0$，而实例 7.15 的余集为 $\bar{X}<\bar{x}_0$ 。

最后，根据“等号总是在原假设上”，找出原假设和备择假设。实例 7.13 中 $H_0：\bar{X}=\bar{x}_0$，$H_1：\bar{X}\neq\bar{x}_0$，实为双侧检验；实例 7.14 中 $H_0：\bar{X}\leqslant\bar{x}_0$，$H_1：\bar{X}>\bar{x}_0$ 为右单侧检验；而实例 7.15 中 $H_0：\bar{X}\geqslant\bar{x}_0$，$H_1：\bar{X}<\bar{x}_0$，是左单侧检验。

2. 选取和计算检验统计量

在参数假设检验中，如同在参数估计中一样，要借助于样本统计量进行统计推断。在具体问题中如何选统计量，需要考虑一些因素，如用于检验的样本是大样本还是小样本，总体方差是已知还是未知等。

对总体均值的假设检验，在大样本($n\geqslant30$)情况下，用 z 统计量，即 $z=\dfrac{\bar{x}-\bar{x}_0}{\dfrac{\sigma}{\sqrt{n}}}$；

在小样本($n<30$)情况下，用 t 统计量，即 $t=\dfrac{\bar{x}-\overline{x_0}}{\dfrac{s}{\sqrt{n}}}$。不同情况下选用不同统计量，这些在后面具体分析。

3. 根据显著性水平查临界值

由于假设检验是利用样本的信息对总体进行推断，就有可能犯错误。例如，原假设正确，而我们却把它当成错误的加以拒绝。犯这种错误的概率用 α 表示，α 称为假设检验中的显著水平(level of significance)。显著水平是指当原假设正确时，人们却把它拒绝的概率。这个概率是由具体问题确定的，通常取 $\alpha=0.05$ 或 $\alpha=0.01$。这表明，当做出接受原假设的决定时，其正确的可能性(概率)为95%或99%。

α 为显著水平，其所对应的概率度称为显著性水平 α 的临界值。例如，$\alpha=0.05$ 时，在正态分布情况下，可以直接利用概率表查找临界值作为判断依据。若为双侧检验，$-Z_{0.025}=-1.96$，$Z_{0.025}=1.96$ 为两个临界值，当 Z 统计量值在［-1.96，1.96］内时，接受原假设；若为右单侧检验，则 $\alpha=0.05$ 查表时，其临界值相当于双侧检验中 $\alpha=0.1$ 所对应的右临界值点，应为 $Z_{0.05}=1.645$，当 Z 统计量值小于等于1.645时，接受原假设；若为左单侧检验，其临界值为 $-Z_{0.05}=-1.645$，当 Z 统计量值大于等于 -1.645 时，接受原假设。

4. 进行比较并做出决策

用计算出的检验统计量的值与临界值相比较，可以做出接受或拒绝原假设的结论，进而可以对所检验的实际问题进行决策。

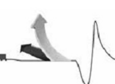
7.4.3 假设检验中的两类错误

假设检验可能犯两种类型的错误。第一类错误(type Ⅰ error)指的是 H_0 客观上真实但被检验所拒绝，这种错误也称为弃真错误。犯这种错误的概率就是显著水平 α。第二类错误(type Ⅱ error)指的是 H_0 客观上不真实但被检验所接受，这种错误也称为纳伪错误。犯这种错误的概率我们用 β 表示。表7-4列出不同实际情况，不同决策对应的4种可能结果的概率。

表7-4 两类错误

	H_0 真实	H_0 不真实
接受 H_0	正确决策($1-\alpha$)	第二类错误(β)
拒绝 H_0	第一类错误(α)	正确决策($1-\beta$)

人们自然希望犯两种错误的概率越小越好，但这是不可能的，因为两者是矛盾的。为使 α 变小就要尽量接受 H_0，其结果必然导致 β 加大。对此，假设检验中一般是保证 α 小到一定标准前提下，再考虑尽量缩小 β，即优先考虑控制“弃真”风险。常用的标准是 $\alpha=0.05$、0.025、0.01，在经济领域中常用 $\alpha=0.05$。



7.4.4 总体均值和总体成数检验

1. 大样本情况下总体均值检验

在大样本($n \geqslant 30$)情况下，总体均值的假设检验可以应用正态分布检验法。

例如，对实例 7.13 以 $\alpha=0.05$ 的显著水平进行假设检验。

解：

第一步，提出原假设和备择假设。$H_0: \overline{X}=227$，$H_1: \overline{X} \neq 227$

第二步，选取和计算检验统计量。$z=\dfrac{\bar{x}-\overline{x_0}}{\dfrac{\sigma}{\sqrt{n}}}=\dfrac{228-227}{\dfrac{3}{\sqrt{100}}} \approx 3.33$

第三步，根据显著性水平查临界值。由 $\alpha=0.05$，得临界值 $-z_{0.025}=-1.96$，$z_{0.025}=1.96$

第四步，进行比较并做出决策。$\because z \approx 3.33 > z_{0.025}=1.96$

$\therefore$ 拒绝 H_0

即这批罐头不符合标准。

而对实例 7.14 进行假设检验($\alpha=0.05$)时应采用单侧检验。

解：

第一步，提出假设。这需要确定检验的方向。题中制造商称轮胎寿命高于 28 000 千米，表述为 $\overline{X}>28\,000$，其余集为 $\overline{X} \leqslant 28\,000$，由于等号在原假设上，故 $\overline{X} \leqslant 28\,000$ 为原假设 H_0，$\overline{X}>28\,000$ 为备择假设 H_1，此为右单侧检验。

第二步，计算统计量。由于总体标准差 σ 未知，用样本标准差 s 替代。

$$\text{统计量 } z=\frac{\bar{x}-\overline{x_0}}{\dfrac{s}{\sqrt{n}}}=\frac{28\,200-28\,000}{\dfrac{1\,000}{\sqrt{40}}} \approx 1.265$$

第三步，查临界值。因为是右单侧检验，$\alpha=0.05$ 时其临界值 $z_{0.05}=1.645$

第四步，比较做决策。$\because z \approx 1.265 < z_{0.05}=1.645$

$\therefore$ 接受 H_0

$\therefore$ 轮胎平均使用寿命不高于 28 000 千米

即制造商广告不属实。

我们再对实例 7.15 以 $\alpha=0.05$ 的显著水平进行假设检验。

解：

第一步，提假设。$H_0: \overline{X} \geqslant 1\,000$ $\quad H_1: \overline{X}<1\,000$

第二步，计算检验统计量。统计量 $z=\dfrac{\bar{x}-\overline{x_0}}{\dfrac{s}{\sqrt{n}}}=\dfrac{960-1\,000}{\dfrac{200}{\sqrt{100}}}=-2$

第三步，查临界值。由 $\alpha=0.05$，左单侧检验得临界值 $-z_{0.05}=-1.645$

第四步，比较做决策。$\because z \approx -2 < -z_{0.05}=-1.645$

$\therefore$ 拒绝 H_0

即批发商不应当购买这批灯泡。

2. 小样本情况下总体均值检验

在小样本($n<30$)情况下，总体均值的假设检验可以应用 t 分布检验法。此时，检验统计量为 t 统计量，$t=\dfrac{\bar{x}-\overline{x_0}}{\frac{s}{\sqrt{n}}}$，它服从自由度为 $n-1$ 的 t 分布。

【实例 7.16】 某食品公司生产袋装食品，采用自动打包机打包。每包标准净重应为 1 000 克，每天开工后需要检验一次打包机工作是否正常。现从产品中随机抽取 9 包，实测每包净重(单位：克)如下：987、993、1 012、1 005、997、983、1 021、995、1 005。给定显著性水平 $\alpha=0.05$，问该日打包机工作是否正常？

解：

由于本题关心的是每包净重是否为 1 000 克，因此是双侧检验问题。又因为是小样本资料，故采用 t 分布检验法。

第一步，提假设。$H_0: \bar{X}=1\ 000$　　$H_1: \bar{X}\neq 1\ 000$

第二步，计算检验统计量。样本平均数 $\bar{x}=999.8$

样本标准差 $s=1.212$

$$t=\frac{\bar{x}-\overline{x_0}}{\frac{s}{\sqrt{n}}}=\frac{999.8-1\ 000}{\frac{1.212}{\sqrt{9}}}=-0.495$$

第三步，查临界值。$\alpha=0.05$，由于是双侧检验，自由度$=n-1=8$，查 t-分布临界值表(见附录 B)，得临界值 $-t_{0.025}(8)=-2.306$，$t_{0.025}(8)=2.306$

第四步，比较做决策。$\because -t_{0.025}(8)=-2.306<t=-0.495<t_{0.025}(8)=2.306$

$\therefore$ 接受 H_0

即该日打包机工作正常。

3. 总体成数检验

在统计实际工作中，经常需要检验总体单位中含有某种特征的单位数所占的比例是否为某个假设值 p_0。例如，一批产品中的合格率，在一次民意测验中表示赞同的比例，等等。此时通常采用大样本资料，因而可以用 z 统计量进行检验。方法及步骤与总体均值检验一样，只是在统计量公式中需要用到成数的标准差。其统计量公式为

$$z=\frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{7-50}$$

【实例 7.17】 一家杂志声称其读者中至少有 25%是在校大学生。从该杂志读者中抽取 200 名读者进行调查，发现其中 42 人是在校大学生。请问该杂志的论断是否正确？($\alpha=0.05$)

解：

第一步，提假设。H_0：$p\geqslant 25\%$，H_1：$p<25\%$

第二步，计算检验统计量。$p=n_1/n=42/200=21\%$



$$z=\frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}=\frac{21\%-25\%}{\sqrt{\frac{25\%(1-25\%)}{200}}}\approx-1.31$$

第三步，查临界值。由 $\alpha=0.05$ 左单侧检验得临界值 $-z_{0.05}=-1.645$

第四步，比较做决策。$\because z\approx-1.31>-z_{0.05}=-1.645$

$\therefore$ 接受 H_0

即该杂志的论断正确。

本章小结

本章共分4节，讲述两类内容。前3节通过讲述抽样推断的基本原理，计算抽样误差，学会如何进行参数估计及如何进行抽样；第4节通过讲述假设检验的基本理论和步骤，学会如何进行总体均值和成数的假设检验。

当需要对总体进行抽样时，应该了解抽样方法有重复抽样和不重复抽样两种；抽样的组织形式共有5种，它们是简单随机抽样、分层抽样、等距抽样、整群抽样和多阶段抽样；抽取的样本容量的确定，分别有平均数计算公式和成数计算公式两种。

（1）平均数的必要样本容量的计算公式。

重复抽样：$n=\frac{t^2\sigma^2}{\Delta^2\bar{x}}$；不重复抽样：$n=\frac{Nt^2\sigma^2}{N\Delta^2\bar{x}+t^2\sigma^2}$

（2）成数的必要样本容量的计算公式。

重复抽样：$n=\frac{t^2P(1-P)}{\Delta p^2}$；不重复抽样：$n=\frac{Nt^2P(1-P)}{N\Delta p^2+t^2P(1-P)}$

当我们需要利用样本资料对总体特征进行推断时，有参数估计和参数假设检验两种方法。

参数估计的一般方法有点估计和区间估计。其中，区间估计是在一定的概率把握程度下，根据样本指标和抽样极限误差估计总体指标所在可能范围的方法。总体平均数和成数的区间估计的公式：$\bar{x}-t\mu_{\bar{x}}\leqslant\bar{X}\leqslant\bar{x}+t\mu_{\bar{x}}$；$p-t\mu_p\leqslant P\leqslant p+t\mu_p$。在进行参数估计时，首先需要求出样本指标 $\bar{x}$ 和 p，然后根据把握程度(可靠度)查表得出对应的概率度 t，最后重要的是求出抽样平均误差 $\mu_{\bar{x}}$ 和 m_p。由于抽样方法有重复抽样和不重复抽样两种，所以不同方法抽取的样本个数不相同。因而计算出的关于平均数和成数的抽样平均误差，均有重复和不重复之分。

假设检验是利用样本的实际资料，检验事先对总体某些数量特征所做的假设是否可信的一种统计分析方法。进行假设检验可能犯两种类型的错误。第一类错误(弃真错误)和第二类错误(纳伪错误)。假设检验的基本步骤：第一步，提假设；第二步，计算检验统计量；第三步，根据显著水平 α 查临界值，第四步，比较做决策。在对总体均值进行假设检验时，若为大样本($n\geqslant30$)资料，应用正态分布检验法，检验统计量为 $z=\frac{\bar{x}-\bar{x}_0}{\frac{\sigma}{\sqrt{n}}}$；若为

小样本($n<30$)资料，采用 t 分布检验法，检验统计量为 $t=\dfrac{\bar{x}-\overline{x_0}}{\dfrac{s}{\sqrt{n}}}$。而对总体成数进行假设检验，一般采用正态分布检验法，检验统计量为 $z=\dfrac{p-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$。

应用与拓展

人的高矮本无优劣，但总有好事者喜欢独辟蹊径，将历史与伟人的身高联系起来考证。一位美国学者对美国总统的身高感兴趣，因为根据资料，美国男人平均身高176厘米，而据美国《华盛顿观察》周刊报道，纵观美国历届总统身高，有24位总统身高都在1.80米以上。在华盛顿政治圈里也盛行“身高论”。在总统竞选中，似乎身高越高的人战胜对手的概率越大。美国总统的身高与普通美国男人身高是否一致？他们之间真有差异吗？

该学者调查了20世纪美国17位总统的身高信息，得到资料见表7-5。

表7-5　20世纪美国17位总统身高资料

序 号	姓　　名	在位年份	身高(美制)/英寸
1	西奥多·罗斯福	1901—1909	70
2	威廉·霍华德·塔夫脱	1909—1913	71.5
3	伍德罗·威尔逊	1913—1921	71
4	沃伦·加梅利尔·哈定	1921—1923	72
5	卡尔文·柯立芝	1923—1929	70
6	赫伯特·胡佛	1929—1933	71.5
7	富兰克林·罗斯福	1933—1945	74
8	哈里·S. 杜鲁门	1945—1953	69
9	德怀特·艾森豪威尔	1953—1961	70.5
10	约翰·肯尼迪	1961—1963	72
11	林登·约翰逊	1963—1969	75.5
12	理查德·尼克松	1969—1974	71.5
13	杰拉尔德·鲁道夫·福特	1974—1977	72
14	吉米·卡特	1977—1981	69
15	罗纳德·里根	1981—1989	73
16	乔治·布什	1989—1993	74
17	比尔·克林顿	1993—2001	74.5

上述资料中的身高是美制单位的“英寸”，将其转换为国际单位“厘米”(1英寸＝2.539 999 918厘米)后，得出20世纪美国17位总统的平均身高 $\overline{X}\approx180.6388$ 厘米，标准差 $s\approx8.5846$ 厘米。利用统计假设检验的方法进行分析如下：



第一步，提假设。$H_0: \bar{X}=176$ $H_1: \bar{X}\neq 176$

第二步，计算检验统计量。样本平均数 $\bar{x}\approx 180.638\,8$

样本标准差 $s=8.584\,6$

$$t=\frac{\bar{x}-\bar{x_0}}{\frac{s}{\sqrt{n}}}=\frac{180.638\,8-176}{\frac{8.584\,6}{\sqrt{17}}}\approx 2.228\,0$$

第三步，查临界值。假如 $\alpha=0.05$，由于是双侧检验，自由度 $=n-1=16$，查 t-分布临界值表，得临界值 $-t_{0.025}(16)=-2.119\,9$，$t_{0.025}(16)=2.119\,9$

第四步，比较做决策。$\because t=2.228\,0>t_{0.025}(16)=2.119\,9$

$\therefore$ 拒绝 H_0

即美国总统身高与美国普通男人身高不一致。

事实上，喜欢夸大身高与“成功”关系者大有人在。2009 年，《高处不胜寒：至高点的生命礼赞》(*The Tall Book: A Celebration of Life from on High*)一书的作者埃尔瑞娜·科恩(Arianne Cohen)接受采访。本身也是高个子(1.90 米)的她说：“我的整个人生都在‘高海拔’中度过，但是没有一本书是讲高个子人生活和工作经历的。所以我想通过写这本书，让更多的人了解我们高个子人的生活经历和心理活动。为什么人会长的这么高等等。这也是一本讲述高个子人成功经历的书。”

实际上，埃尔瑞娜个人独特的经历本身就具有代表性。在哈佛大学学习生涯中，她选修了 35 门课程，跨越了 29 个系，以至于她在大学三年级时就被人认为无法毕业，因为缺少专业。然而，她不仅获得了拉丁文学位，还以极优等水平毕业。目前，她在《纽约时报》、时尚杂志《嘉人》、*vogue*、《科技新时代》等杂志拥有固定专栏。

早在几年前，哈佛大学著名经济学教授格雷戈里·曼昆(Gregory Mankiw)和马修(Matthew)提出应对高个子人征收更多的税，因为他们赚的钱更多。对此，埃尔瑞娜认为这是一种无礼行为。“幸运的是，国会议员中有一半以上都是 1.90 米以上的高个子。”她说。

(资料来源：http://news.fznews.com.cn/guoji/2009-6-17/2009617ic5ob7hv_17213.shtml.)



习题与实训

一、单项选择题

1. 下面不属于抽样推断的特点的是(　　)。

A. 抽样推断是建立在随机取样的基础上

B. 抽样推断是运用概率估计的方法，利用样本指标估计总体参数

C. 抽样推断是由部分对整体总量进行的概括估计而非精确推断

D. 抽样推断的误差是可以预先控制的

2. “方差”是(　　)。

A. 离差的平方　　B. 平均差的平方

C. 标准差的平方　　D. 变异系数的平方

3. 在同等条件下，重复抽样与不重复抽样相比较，其抽样平均误差(　　)。

A. 前者小于后者　　B. 前者大于后者

C. 两者相等　　D. 无法确定哪一个大

4. 抽样推断中的概率保证程度表达了区间估计的(　　)。

A. 显著性　　B. 准确性　　C. 可靠性　　D. 规律性

5. 抽样极限误差和抽样平均误差之间的关系为(　　)。

A. 极限误差可以大于、等于或小于平均误差

B. 极限误差一定大于平均误差

C. 极限误差一定小于平均误差

D. 极限误差一定等于平均误差

6. 置信概率定得越大，则置信区间相应(　　)。

A. 越小　　B. 越大　　C. 变小　　D. 不变

7. 比较简单随机抽样情况下重复与不重复抽样误差的大小，如果在总体中按不重复抽样抽取 36%做样本，则抽样误差比重复抽样的误差要小(　　)。

A. 8%　　B. 20%　　C. 36%　　D. 64%

8. 在右侧检验中，H_0：$M \leqslant M_0$，则拒绝域为(　　)。

A. $|t| \geqslant t_\alpha$　　B. $t \leqslant -t_\alpha$　　C. $t \leqslant -t_{\alpha/2}$　　D. $t \geqslant t_\alpha$

9. 对显著水平 α 的描述错误的是(　　)。

A. 当做出接受原假设的决定时，其正确的可能性(概率)为 $1-\alpha$

B. 显著水平是当原假设正确时，人们却把它拒绝的概率

C. α 是犯第二类错误的概率

D. 显著水平越大，犯纳伪错误的概率越小

10. 第一类错误与第二类错误的关系是(　　)。

A. 两类错误同时增加，样本容量不变

B. 两类错误同时减小，样本容量增大

C. 两类错误同时减小，样本容量不变

D. 两类错误同时增加，样本容量增大

二、多项选择题

1. 抽样调查的特点是(　　)。

A. 按随机原则抽取样本　　B. 按随意原则抽取样本

C. 由部分推断总体　　D. 可以事先计算并控制抽样误差

E. 缺乏科学性和可靠性

2. 按组织方式不同，抽样调查有(　　)。

A. 简单随机抽样　　B. 等距抽样　　C. 类型抽样

D. 整群抽样　　E. 重复抽样和不重复抽样

3. 抽样调查中的抽样误差是(　　)。

A. 不可避免的　　B. 可以事先计算并加以控制

C. 抽样估计值与总体参数值之差　　D. 可以避免的

E. 受总体标志变动程度的影响



4. 影响抽样误差的主要因素有(　　)。

A. 抽样数目的多少　　B. 总体标志变异程度的大小
C. 不同的组织方式　　D. 抽样周期的长短
E. 不同的抽样方法

5. 要提高抽样推断的精确度，可采用的方法有(　　)。

A. 增加样本数目　　B. 减少样本数目
C. 缩小总体被研究标志的变异程度　　D. 改善抽样的组织方式
E. 改善抽样的方法

6. 影响抽样数目的主要因素有(　　)。

A. 总体被研究标志的变异程度大小　　B. 抽样的组织方式
C. 对推断精确度的要求　　D. 对推断把握程度的要求
E. 抽取调查单位的方法

7. 在简单重复随机抽样条件下，欲使误差范围缩小$\frac{1}{2}$，其他要求保持不变，则样本容量必须(　　)。

A. 比原来增加2倍　　B. 比原来增加3倍
C. 是原来的4倍　　D. 比原来减少2倍
E. 是原来的3倍

8. 一个完整的假设检验过程，通常包括的步骤为(　　)。

A. 提出原假设和备择假设　　B. 选取和计算检验统计量
C. 计算抽样平均误差和抽样极限误差　　D. 根据显著性水平查临界值
E. 进行比较并做出决策

9. 关于原假设与备择假设的叙述正确的有(　　)。

A. 把需要通过样本推断其正确与否的命题称为原假设，用H_0表示
B. 备择假设是与原假设相对立的假设
C. 原假设与备择假设一定是互斥的，没有相交内容
D. 等号总是在原假设上
E. 等号可能在原假设上，也可能在备择假设上

10. 若要检验“大学生初职平均预期月薪是否高于3 000元”，则(　　)。

A. 采用双侧检验　　B. 采用单侧检验
C. 采用左单侧检验　　D. 采用右单侧检验
E. 左单侧和右单侧检验均可

三、判断题

1. 在总体方差一定的条件下，样本单位数越多，则抽样平均误差越大。　　(　　)

2. 从总体N个单位中，用重复抽样的方法，随机抽样取n个单位构成一个样本，则共可抽取N^n个样本。　　(　　)

3. 在其他条件不变的情况下，提高抽样估计的可靠程度，可以提高抽样估计的精确度。　　(　　)

4. 抽样极限误差总是大于抽样平均误差。（ ）

5. 若针对总体特征提出的假设为 H_0：$\overline{X}=\bar{x}_0$，H_1：$\overline{X}\neq\bar{x}_0$，则将采用双侧检验。（ ）

6. 概率保证程度与概率度 t 密切联系，当 t 为 2 时，概率保证程度为 95%。（ ）

7. 在简单重复随机抽样中，将抽样极限误差缩小 1/3，则样本容量需扩大 1.15 倍。（ ）

8. 在正态分布情况下 $\alpha=0.05$ 时，若为双侧检验，有两个临界值分别是 $-z_{0.025}=-1.96$，$z_{0.025}=1.96$。（ ）

9. 如果要检验“某产品的平均重量是否显著小于 454 克”，应采用右侧检验。（ ）

10. 在简单重复随机抽样中，其他条件保持不变，只将抽样的可靠程度由 68.27%提高 27.18 个百分点，则样本容量须由原来的 100 个变成现在的 400 个。（ ）

四、填空题

1. 抽样推断是在____________的基础上，利用样本资料计算样本指标，并据以推算________特征的一种统计分析方法。

2. 假定 10 亿人大国和 100 万人小国的居民年龄变异程度相同。现各自用重复抽样的方法抽取本国的 1%人口计算平均年龄，则平均年龄抽样平均误差比较小的是________。

3. 重复抽样情况下，若其他条件不变，将概率保证程度由 95%提高到 95.45%，同时抽样极限误差缩小 1%，则样本容量将是原来的________倍。

4. 两个总体标志变异度不同，甲总体标准差是乙总体标准差的 1.1 倍，现从两总体抽样，抽样的精确度和可靠度一样，但抽取样本的数目甲是乙的________。

5. 对产品合格率进行抽样检查时，若规定极限误差不超过 3%，概率保证程度为 95%，则最多需抽取的样本个数是________。

6. 为了估计某餐馆每位顾客午餐的平均花费，在为期 3 周的时间里选取 49 名顾客进行抽样调查，得到样本均值为 22.6 元，标准差为 2.5 元，则总体平均花费的 95%的置信区间是________。

7. 某报社宣称其报纸覆盖率不低于当地人群的 30%，为验证这一声明是否属实，对抽样结果进行检验，这一假设检验应采用________。

8. 抽样估计的方法有两种，即________和________。

9. 某学校随机抽查 10 名男学生，测得平均身高 170 厘米，标准差 15 厘米，若估计全校男学生身高介于 160.5～179.5 厘米，则其把握程度是________。

10. 抽样调查过程中可能发生的误差分为________和________两类。

五、应用能力训练题

1. 某市电视台在北京奥运会期间开辟制作了一专栏节目，在该市随机对 900 名居民进行调查，结果有 540 名喜欢该电视栏目，则

（1）此次抽样误差是多少？

（2）若置信度为 95%，那么抽样的最大极限误差是多少？

（3）以 95%概率保证程度，估计该市居民喜欢该电视栏目的比率范围。



(4) 若某人估计该市居民喜欢该电视栏目的比率在57.3%～62.7%，则把握程度大约为多少?

2. 对某地一城市进行居民家庭人均旅游消费支出调查，随机抽取400户居民家庭，调查得知居民家庭人均年旅游消费支出为400元，标准差为100元，则以95%概率推断该市居民家庭人均年旅游消费支出区间范围。

3. 平安保险公司利用由36个投保人所组成的简单随机样本来估计全体投保人的平均年龄，其精确度建立在0.95的概率保证下，抽样平均数的抽样极限误差要小于或等于2.35岁的基础上。已知总体标准差为7.2岁，假定该保险公司根据工作需要，希望提高精确度，使抽样极限误差降低至小于或等于2岁，那么需要多大的样本容量?

4. 长平化工厂有1 500名工人，用简单随机重复抽样的方法抽出50名工人作为样本，调查其工资水平，资料见表7-6。

表7-6　被调查的50名工人的工资水平资料

月平均工资/元	3 144	3 204	3 240	3 300	3 360	3 480	3 600	3 960
工人人数/人	4	6	9	10	8	6	4	3

根据以上资料：

(1) 计算样本平均数和抽样平均误差。

(2) 以95.45%的概率估计该厂工人的月平均工资和工资总额的区间。

5. 从一批出口产品中按不重复随机抽样方法抽选200件，检测出废品8件，又知道抽样数是该批产品总数的1/20，当概率为0.954 5时，可否认为这批产品的废品率不超过5%?

6. 现有一批袋装食品共10万袋，用重复抽样方法，检查其每袋平均重量及产品合格率。规定平均每袋重量的误差范围不超过0.2克，合格率的允许误差不超过12%，据以往资料，产品每袋重量的方差为3.5，置信度为99.73%，则应抽查多少袋?

7. 某砖厂对所生产的产品进行质量检查，要求允许误差不超过0.011 11，把握程度为0.954 5，并知道过去进行几次同样调查所得合格产品百分比为98.75%和98.88%，那么必要抽样单位数目是多少?

8. 外贸公司出口一种食品，规定每包规格不低于150克，现在用不重复抽样的方法抽取其中的100包进行检验，其结果见表7-7。

表7-7　被检验食品的抽样结果

每包重量/克	包数/包	每包重量/克	包数/包
148～149	10	151～152	20
149～150	20	合计	100
150～151	50		

根据以上资料：

(1) 以99.73%概率估计这批食品平均每包重量的范围，以便确定平均重量是否达到规格要求。

(2) 以同样的概率估计这批食品合格率范围。

9. 某公司接受调查一种新型洗衣机的销售前景。拟通过抽样调查，了解居民在收听(看)有关广告节目后对新型洗衣机感兴趣的户数比例。据从前推销其他商品的经验，这个比例不会超过20%，若确定置信区间宽度为4个百分点，取置信水平为95%，样本容量是多少?

10. 某学校有2 000名学生参加英语等级考试，为了解学生的考试情况，用不重复抽样方法抽取部分学生进行调查，所得资料见表7-8。

表7-8 被抽取学生的调查资料

考试成绩/分	60以下	60～70	70～80	80以上
学生人数/人	20	20	45	15

试以95.45%的概率估计该校学生英语等级考试成绩在70分以上的学生所占比例范围。

11. 根据国家标准，某食品的色素含量不得超过0.36克。一家食品厂宣称，其食品的色素含量符合国家这一标准。食品检验部门从该厂的产品中随机抽取20个样品，测得平均色素含量为0.38克，标准差为0.05克。在5%的显著性水平下，该厂生产的食品是否达到了国家标准?

12. 某企业进口一批金属板材，合同约定的板材厚度为0.8厘米。现随机抽取64块板材，测得其厚度，计算得出样本平均厚度为0.78厘米，样本标准差为0.024厘米。检验判断这批金属板材平均厚度是否明显低于合同约定的板材厚度。($\alpha=0.05$)

13. 某制造厂生产某装置的平均工作温度是190℃。现从一个由16台装置构成的随机样本求得的工作温度的平均数和标准差分别是194℃和8℃，设$\alpha=0.05$，并假定工作温度服从正态分布，请检验"平均工作温度是否比制造厂规定的要高"。

14. 某种大量生产的袋装食品，按规定不得少于250克，现从一批该食品中任意抽取50袋，发现有6袋低于250克，若规定不符合标准的比例超过5%就不得出厂，显著性水平为0.05，则这批食品能否出厂?

15. 某市居民中平常订阅某种报纸的占40%，但从2008年第四季度看，订阅率似乎出现减少的现象，于是随机抽取200户家庭进行调查，发现有76户家庭订阅该报纸，则以$\alpha=0.05$的显著性水平推断，该报纸的订阅率与以往相比是否显著降低。

16. 某房产商宣称某邻近地区房屋每间平均价格大于45 000元，现以其36间房屋组成一个随机样本，得出的平均价格为48 000元，均方差为12 000元，则在0.05显著性水平下，这些数据是否支持该房产商的说法?

17. 某大学在2013年5月底进行了一次关于大学生吸烟问题的问卷调查，回收问卷842份，其中有101人声称自己吸烟，而根据资料显示，当地城市居民中20%人吸烟，则大学生中吸烟者比例是否高于当地居民吸烟比例的1/2?($\alpha=0.05$)



第8章 相关分析与回归分析

理论目标

(1) 理解相关关系的概念与种类。
(2) 掌握相关系数的性质和理论。
(3) 掌握回归系数的含义及计算。
(4) 掌握相关分析与回归分析的基本思路和步骤。

能力目标

(1) 能利用相关图表和相关系数判断现象是否具有相关关系及相关方向和程度。
(2) 能够确定两个相关变量的简单线性回归方程。
(3) 能根据回归方程进行预测。

做一做

广告的投入与产品的销售之间的关系世人皆知。然而，在"广告投入越多，产品销售越好"共性下，不同行业不同产品的"广告投入效率"却各有差异。王强从事啤酒销售工作，在2012年他收集了国内10个品牌啤酒的广告费用和销售量的数据，见表8-1。

表8-1 2012年国内10个品牌啤酒的广告费用和销售量

广告费用/万元	120	68.7	100.1	76.6	8.7	1	21.5	1.4	5.3	1.7
销售量/万箱	36.3	20.7	15.9	13.2	8.1	7.1	5.6	4.4	4.4	4.3

根据以上材料，王强能否判断广告费用与销售量之间有相关关系？若有关系，关系程度如何？如果

关系程度紧密，可用怎样的表达式近似表达两者的量化关系？按此规律，广告费用每增加10 000元，销售量将有如何变化？

想一想

凭经验知道广告费用与销售量有关系，那么如何形容它们关系的密切程度呢？如果用一个具体的量化指标来反映两个现象之间的相关方向和程度，那对问题就很容易清晰地分析了，这个指标是如何定义的，怎样计算？进一步讲，若已知两个变量间关系紧密，一个变量变动，另一个必然相应发生变化，问题是一个变量变动一个单位，另一个变量增加或减少几个单位？亦即希望知道两现象间量化关系，其近似函数表达式是什么。因为如果有了这个表达式，可以通过给出的一个变量值来预测另一个变量的数值，这将对我们的经济生活很有意义。要想探究经济生活中这些问题，就需要引进新的统计工具——相关分析和回归分析。

8.1 相关分析

8.1.1 相关关系的概念

相关关系(correlation)是指现象间的非确定性的数量上的依存关系。它有两个特点：一是现象之间确实存在数量上的依存关系；二是数量依存关系的值是不确定的。例如，人的身高和体重之间确实存在着“身高越高，体重越大”的数量关系，但是，身高和体重却不是一一对应的。身高165cm的人，相应的体重并不完全一样；同样，体重65kg的人，会对应着不同的身高。相关关系的这种特点，决定了它与函数关系的区别。

函数关系是指现象间存在的确定性的数量依存关系。在这种关系中，某个现象的数值发生变化，都有另一现象的确定值与它对应。这种关系可用数学函数式反映。例如，给定一个圆的半径，就有唯一确定的圆的面积和它对应，面积是半径的函数。在社会经济中，产品生产费用(总成本)=产品产量×单位产品成本，当单位产品成本不变时，产品产量发生变化，就有一个确定的总成本与它对应，总成本是产品产量的函数。

函数关系与相关关系既有区别又有联系。由于观察和实验中的误差，函数关系往往通过相关关系表现出来。而当对现象之间的内在联系和规律性了解得更加清楚的时候，相关关系有可能转化为函数关系。一般来说，在社会经济领域里，函数关系反映了现象之间的理想状态，而相关关系则反映了现象之间的现实状态。在对具有相关关系的现象进行分析时，必须利用相应的函数关系数学表达式来表明现象之间的相关方程式。相关关系是相关分析(correlation analysis)的研究对象，函数关系是相关分析的工具。

8.1.2 相关关系的种类

1. 按相关程度区分

按相关程度不同，相关关系可分为完全相关、不相关和不完全相关。

两种现象之间，若当一个现象的数量确定时，另一现象的数量也随之确定，则这两种



现象间的关系称为完全相关，其关系如图 8.1 所示。例如，某年中国银行的 1 年期居民储蓄存款利率为 3.25%，存入的本金用 x 表示，到期的本利和用 y 表示，则 $y=x+3.25\%x$。这里 y 与 x 表现为一种线性函数关系，也就是说函数关系是相关关系的一个特例。

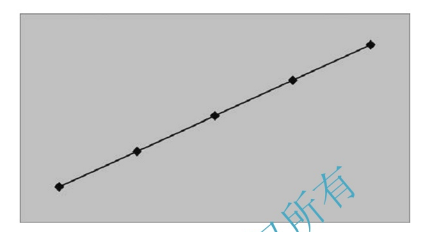


图 8.1 完全相关

如果两个现象之间的数量变化相互独立，彼此互不影响，则这两种现象间的关系称为不相关，其关系如图 8.2 所示。例如，学生的学习成绩与身高之间不存在相关关系。

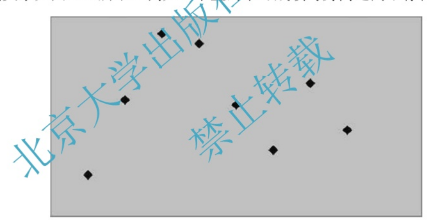


图 8.2 不相关

如果两个现象之间的关系介于完全相关和不相关之间，则称为不完全相关，其关系如图 8.3 所示。例如，产品的销售量与广告费用之间是不完全相关关系。现实经济生活中大多数相关关系都属于不完全相关，这是统计分析的主要研究对象。

2. 按相关方向区分

按相关方向不同，相关关系可分为正相关和负相关。

相关变量的变化方向一致，即一个变量数值增大(或减少)，另一个变量的数值也相应地增大(或减少)，这种相关关系称为正相关，如图 8.3 所示。例如，广告费用与销售量、居民的收入与居民的储蓄额的关系等，都属于正相关。

若相关变量的变化方向相反，即一个变量数值增大(或减少)，另一个变量的数值相应地减少（或增大)，这种相关关系称为负相关，如图 8.4 所示。例如，家庭收入增加，恩格尔系数下降，它们之间属于负相关关系。

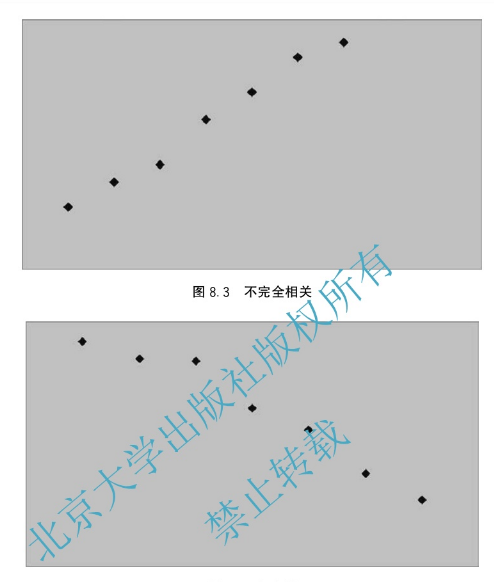


图 8.3　不完全相关

图 8.4　负相关

3. 按相关表现形式区分

按相关表现形式不同，相关关系可分为线性相关和非线性相关。

对两个具有相关关系的现象进行实际调查，能获得一系列成对的数据。每对数据在平面直角坐标系中确定一个点，如果这些点大致分布在一条直线周围，则这两种现象就构成线性相关，如图 8.3 和图 8.4 所示。例如，在一定范围内，出租汽车费用与行驶里程呈线性相关关系；若这些点的分布并不表现为直线关系，而是近似于某种曲线关系，则这种关系就称为非线性相关，如图 8.5 所示。例如，从人的生命过程来看，年龄与医疗费支出呈非线性相关。

4. 按相关变量多少区分

按相关变量多少不同，相关关系可分为单相关和复相关。

两个变量之间的相关关系称为单相关，这种相关关系中只涉及一个自变量和一个因变



量，如图 8.1～图 8.5 所示。例如，广告费用和啤酒的销售量的关系就是单相关。

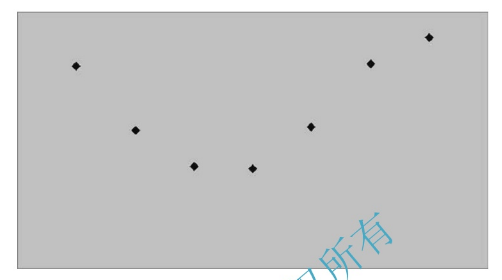


图 8.5 非线性相关

3 个或 3 个以上变量的相关关系称为复相关，这种相关关系中涉及两个或两个以上自变量和一个因变量。例如，冰淇淋的销售量与价格、气温、居民收入，以及居民年龄构成等之间的相关关系就是复相关。

统计学中对相关关系进行分析的主要目的，是要对现象之间的相互关系的密切程度和变化规律进行分析，取得具体数量上的认识，并进一步确定出相互关系的模式，以便进行统计预测和推算。根据对现象分析研究层次的不同，相关分析主要内容有狭义与广义之分。

狭义地讲，相关分析主要包括以下内容。

(1) 判断现象之间是否存在相关关系(通过编制相关表和绘制相关图)。

(2) 确定相关关系的密切程度和方向(通过计算相关系数)。

广义地说，相关分析主要内容除了包括狭义相关分析内容外，还包括以下内容。

(1) 确定相关关系的数学表达式(通过建立回归方程)。

(2) 确定因变量估计值误差的程度(通过计算估计标准误差)。

下面逐项进行研究。

8.1.3 相关图表

相关分析的起点是定性分析，按照认识的逻辑顺序分析并判断现象之间是否存在相关关系。在此基础上，进一步进行定量分析，编制相关表(correlation table)，绘制相关图(correlation diagram)，利用相关图表判断相关关系的形式，为相关分析奠定基础。

1. 相关表

相关表是一种反映变量之间相关关系的统计表。将其中一个变量按其取值大小顺序排列，再将与其相关的另一个变量数值对应列出来，便可形成简单相关表。通过相关表，可初步看出相关关系的形式、密切程度和相关方向。现举例说明。

【实例 8.1】 根据“做一做”中表 8-1 的资料，判断两者的相关关系。

解：

根据表 8-1 的原始资料，将广告费用按从小到大顺序排列，可编制简单相关表，见表 8-2。

表 8-2 2012 年国内 10 个品牌啤酒的广告费用和销售量

广告费用/万元	1	1.4	1.7	5.3	8.7	21.5	68.7	76.6	100.1	120
销售量/万箱	7.1	4.4	4.3	4.4	8.1	5.6	20.7	13.2	15.9	36.3

从表 8-2 可以看出，随着广告费用的提高，啤酒销售量也有相应提高的趋势，两者之间存在着明显的正相关关系。

2. 相关图

相关图又称散点图(scatter plot)，它是把相关表中的原始对应数据在平面直角坐标系中以坐标点描绘出来。以横轴表示自变量(dependent variable)，纵轴表示因变量(independent variable)，通过标出的每对变量值的坐标点或散布点，观察其分布状况。例如，可将表 8-1 中所列的资料绘制成相关图，如图 8.6 所示。

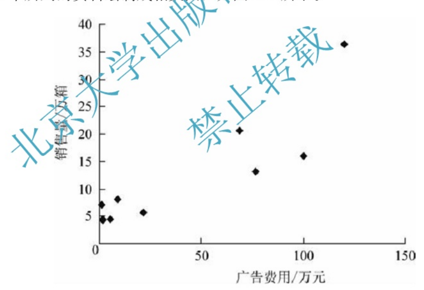


图 8.6 广告费用与销售量的相关关系

从图 8.6 中关于销售量及广告费用的 10 个散布点看，销售量随着广告费用的增大而增大，随之减小而减小，并且散布点的分布近似地表现为一条直线。由此可以判断，啤酒销售量与广告费用这两个变量之间存在着直线正相关关系。

但是，相关表和相关图只能大体上反映现象之间的相关关系，为了更为精确地反映关系类型的特征，还应该进一步用统计分析指标——相关系数来表明相关的密切程度。按照相关变量的个数和分析问题的角度不同，相关系数可以分为简单相关系数、偏相关系数和复相关系数。本章只讨论简单相关系数，它适用于影响因变量的因素只有一个自变量的情况。



8.1.4　相关系数

1. 相关系数的意义

相关系数(correlation coefficient)是在直线相关条件下，说明两个变量之间相关关系密切程度和方向的统计分析指标，通常用 r 表示。其表现形式为相对数，它不受变量值水平和计量单位的影响。

相关系数的取值有一个范围，$-1 \leqslant r \leqslant 1$；有两个方向，$r>0$ 表示正相关，$r<0$ 表示负相关；有 3 个特殊点，$r=1$ 表示完全正线性相关，$r=-1$ 表示完全负线性相关，$r=0$ 表示完全不线性相关；有 4 个相关密切程度等级，$0<|r|<0.3$ 为微相关，$0.3 \leqslant |r| <0.5$ 为低度相关，$0.5 \leqslant |r| <0.8$ 为显著相关，$0.8 \leqslant |r| <1$ 为高度相关。

采用上述标准进行判断时，需要注意 r 只表示 x 和 y 的直线相关密切程度，当 $|r|$ 很小甚至等于 0 时，并不一定表示 x 与 y 之间就不存在其他非直线类型的关系。另外，要注意应该用较多原始数据计算相关系数。当原始资料较少时，为了判断得准确，最好对相关系数的显著性进行检验，关于检验方法，本书不进行讨论。

2. 相关系数的计算

在直线相关的条件下，相关系数的定义公式是通过自变量和因变量的各个离差的乘积来表明相关关系的密切程度的，所以用这种方法计算的相关系数叫积差法。

相关系数的基本公式为

$$r=\frac{\sigma_{xy}^{2}}{\sigma_{x}\sigma_{y}} \tag{8-1}$$

式中：$\sigma_{xy}^{2}=\dfrac{\sum(x-\bar{x})(y-\bar{y})}{n}$ 称为协方差；$\sigma_{x}=\sqrt{\dfrac{\sum(x-\bar{x})^{2}}{n}}$ 是 x 的标准差；$\sigma_{y}=\sqrt{\dfrac{\sum(y-\bar{y})^{2}}{n}}$ 是 y 的标准差。

所以相关系数可表现为

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^{2}}\cdot\sqrt{\sum(y-\bar{y})^{2}}} \tag{8-2}$$

这是计算相关系数的基本公式，但在实际工作中利用它来计算相关系数相当烦琐，经代数演算，可化简为较简单的、能直接用相关表中的原始数据计算的计算公式。计算公式为

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^{2}-\left(\sum x\right)^{2}}\cdot\sqrt{n\sum y^{2}-\left(\sum y\right)^{2}}} \tag{8-3}$$

式中：n 为相关表中数据项数。

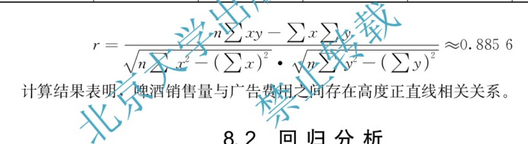
【实例 8.2】　根据表 8-1 的资料，已知啤酒销售量与广告费用之间为直线相关关系，计算啤酒销售量与广告费用的相关系数。

解：

根据表 8－1 的数据，可以编制啤酒销售量与广告费用的相关系数计算表，见表 8－3。

表 8－3　相关系数计算

编　号	广告费用 x/万元	销售量 y/万箱	x^2	y^2	xy
1	120	36.3	14 400	1 317.69	4 356
2	68.7	20.7	4 719.69	428.49	1 422.09
3	100.1	15.9	10 020.01	252.81	1 591.59
4	76.6	13.2	5 867.56	174.24	1 011.12
5	8.7	8.1	75.69	65.61	70.47
6	1	7.1	1	50.41	7.1
7	21.5	5.6	462.25	31.36	120.4
8	1.4	4.4	1.96	19.36	6.16
9	5.3	4.4	28.09	19.36	23.32
10	1.7	4.3	2.89	18.49	7.31
合计	405	120	35 579.14	2 377.82	8 615.56



$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-\left(\sum x\right)^2}\cdot\sqrt{n\sum y^2-\left(\sum y\right)^2}}\approx 0.8856$$

计算结果表明，啤酒销售量与广告费用之间存在高度正直线相关关系。

8.2　回 归 分 析

8.2.1　回归分析的意义

回归(regression)这个术语是由英国著名统计学家弗朗西斯·高尔顿(Francis Galton)于 19 世纪末期在研究孩子及他们的父母的身高时提出来的。高尔顿对 1 078 对父子身高关系进行了研究，假设父亲的身高是影响的因素，记为 x(单位：inch，1inch≈2.54cm)，其儿子的身高是影响的结果，记为 y(单位：inch)。经过对资料进行整理、分组后，确定出那些资料中大约存在如下的线性关系：

$$y_c=33.73+0.516x \tag{8-4}$$

式中：y_c表示假设父亲身高为 x 时，其儿子的身高的估计值。

这就是著名的回归直线方程的雏形。之所以称为回归直线方程，首先它是直线方程，斜率 0.516 表明，父亲身高每增加 1inch，其儿子的身高平均增加大约 0.5inch。所谓回归，是指虽然该方程反映 y_c与 x 是正比的关系，但是，当 x 较小(即父亲较矮)时，y_c(即儿子的身高)可能比 x 大；而当 x 较大时，y_c可能比 x 小，有“回归”之意。高尔顿把这种孩子的身高向中间值靠近的趋势称为一种回归效应，并把“回归”作为统计概念加以应



用。但现今统计学的“回归”概念已不是指原来生物学上的特殊规律性，而是指变量之间的依存关系。

回归分析(regression analysis)是指研究一个或几个变量的变动对另一个变量的变动影响程度的方法。根据资料，建立变量之间的回归数学模型，即建立相应的数学表达式，通过给定的自变量的数值估计因变量的可能值，这种分析就叫回归分析，所建立的数学模型称回归模型。回归模型有多种形式，按自变量个数分，有一元回归(即简单回归，simple regression)和多元回归(即复回归，multiple regression)；按回归线形状分，有线性回归(即直线回归，linear regression)和非线性回归(即曲线回归，nonlinear regression)。其中，线性回归是基本的，本章只介绍一元线性回归，即简单线性回归分析方法。

8.2.2　回归分析的特点

相关分析是回归分析的基础和前提，回归分析是相关分析的深入和继续。

回归分析与相关分析相比具有以下特点。

(1) 在相关分析中两变量是随机变量；而回归分析中，因变量是随机变量，自变量是给定的数值。

(2) 相关分析只能计算出一个反映两个变量间相关密切程度的相关系数，且两个变量是对等关系，变量 x 和 y 位置互换不影响相关系数的数值。而回归分析所研究的两个变量不是对等关系，有时可根据研究目的不同建立两个不同的回归方程，一个是以 x 为自变量，y 为因变量的“y 对 x 的回归方程”；另一个是以 y 为自变量，x 为因变量的“x 对 y 的回归方程”。

(3) 相关分析计算的相关系数是一个绝对值在 0～1 的抽象系数，其数值大小反映变量之间相关关系的程度；而回归分析建立的回归方程反映的是变量值之间的具体变动关系，不是抽象的系数。根据回归方程，利用自变量的给定值，可以估计或推算因变量的数值。

8.2.3　一元线性回归方程

一元线性回归方程(regression equation)，亦称直线方程，是分析一个自变量 x 与一个因变量 y 之间线性关系的数学方程。方程的基本形式为

$$y_c = a + bx \tag{8-5}$$

a 和 b 称为回归方程中的两个待定参数，是需要根据相关表中的 x 与 y 的实际资料求解的数值，a 和 b 的值确定了直线的位置，而它们一旦确定，这条直线就被唯一确定了。但用于描述这 n 组数据的直线有许多条，究竟用哪条直线代表两个变量之间的关系，则需要有一个原则。即希望这条直线离各离散点最近，具体表述为对于相关表中的 x，它对应的实际数值 y 同这条直线上的理论值 y_c 的离差平方和为最小值。用公式表示为

$$\sum (y - y_c)^2 = \sum (y - a - bx)^2 = 最小值 \tag{8-6}$$

设 $Q = \sum (y-a-bx)^2$，则 Q 是两个待定参数 a 和 b 的函数。要使 Q 为最小值，就要用对二元函数求极值的原理，求 Q 关于 a 和 b 的偏导数，并令其等于零。经过整理得出直

线回归方程中求解参数 a 和 b 的标准方程组：

$$\begin{cases}\sum y = na + b\sum x \\ \sum xy = a\sum x + b\sum x^2\end{cases} \tag{8-7}$$

解此方程组，得出求解 a、b 的计算公式为

$$\begin{cases}b = \dfrac{n\sum xy - \sum x\sum y}{n\sum x^2 - (\sum x)^2} \\ a = \dfrac{\sum y}{n} - b\dfrac{\sum x}{n}\end{cases} \tag{8-8}$$

根据这一思想确定未知参数 a 和 b 的方法，称为最小平方法(least squares analysis)。

【实例 8.3】 仍以“做一做”中表 8－1 给出的 2012 年国内 10 个品牌啤酒广告费用及销售量为例，运用最小平方法，求解回归方程参数 a 和 b，并建立一元线性回归模型。

解：

设所配合的一元线性回归方程为 $y_c = a + bx$，其中，设广告费用为 x，销售量为 y。根据表 8－1 资料可得表 8－4 的结果。

表 8－4 直线回归模型计算

编　号	广告费用(x)/万元	销售量(y)/万箱	x^2	xy
1	120	36.3	14 400	4 356
2	68.7	20.7	4 719.69	1 422.09
3	100.1	15.9	10 020.01	1 591.59
4	76.6	13.2	5 867.56	1 011.12
5	8.7	8.1	75.69	70.47
6	1	7.1	1	7.1
7	21.5	5.6	462.25	120.4
8	1.4	4.4	1.96	6.16
9	5.3	4.4	28.09	23.32
10	1.7	4.3	2.89	7.31
合计	405	120	35 579.14	8 615.56

根据最小平方法确定参数的公式，可计算出参数 a 和 b 为

$$\begin{cases}b = \dfrac{n\sum xy - \sum x\sum y}{n\sum x^2 - (\sum x)^2} \\ a = \dfrac{\sum y}{n} - b\dfrac{\sum x}{n}\end{cases}$$



$$b=0.1958$$

$$a=4.07$$

将 a 和 b 代入回归方程，则 $y_c=4.07+0.1958x$

回归直线确定后，将各品牌啤酒的广告费用依次代入方程式，即可求得相应的销售量理论值。

根据所建立的直线回归方程，也可以进行外推预测。例如，广告费用为 110 万元，在其他条件相对稳定时，可以预测其销售量为 $y_c=4.07+0.1958\times110=25.608$(万箱)。

在这里，下列几个问题应引起注意。

(1) 回归方程中，$a=4.07$，是销售量的起点值，即广告费用 $x=0$ 时，y_c的值。

(2) b 称为回归系数(regression coefficient)。$b=0.1958$，表示广告费用每增加一个单位(1 万元)，啤酒销售量平均增加 0.195 8 个单位(万箱)。

(3) 由于 b 的计算公式中的分子与相关系数 r 计算公式中的分子完全一样，且 b 与 r 为正值还是负值，其符号均取决于这个分子，所以，回归系数 b 与相关系数 r 的符号必然一致。即通过回归系数 b 的符号可判断两变量相关的方向。

(4) 回归系数 b 与相关系数 r 之间还存在密切的数量关系，两者可相互推算。根据相关系数 $r=\frac{\sigma_{xy}^2}{\sigma_x\sigma_y}$和回归系数 $b=\frac{n\sum xy-\sum x\sum y}{n\sum x^2-\left(\sum x\right)^2}=\frac{\sigma_{xy}^2}{\sigma_x^2}$，可知 $\frac{b}{r}=\frac{\sigma_y}{\sigma_x}$，亦即 $b=r\cdot\frac{\sigma_y}{\sigma_x}$。

(5) 另外，一个直线回归方程只能做一种推算，不能反向进行推算。也就是说，只能以自变量 x 推算因变量 y，而不能以因变量 y 推算自变量 x。如上例所配合的直线回归方程，只能在给定广告费用数值时推算销售量，不能以给定的销售量数值反过来推算广告费用。

若根据研究需要，想了解当销售量一定时，相对应的广告费用数值应达到多少，则应建立 x 对 y 的直线回归方程，即 $x_c=c+dy$。式中 c 与 d 的意义与公式 $y_c=a+bx$ 中 a、b 意义相同，均为待定参数，只是 x 与 y 的位置互换而已。同样采用最小平方法，确定出参数 c 与 d 公式为

$$\begin{cases} d=\dfrac{n\sum xy-\sum x\sum y}{n\sum y^2-\left(\sum y\right)^2} \\ c=\dfrac{\sum x}{n}-d\,\dfrac{\sum y}{n} \end{cases} \tag{8-9}$$

【实例 8.4】 根据“做一做”中表 8-1 给出的 2012 年国内 10 个品牌啤酒广告费用及销售量，预测若销售量为 30 万箱时，相对应的广告费用数值。

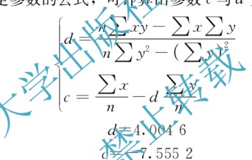
解：

若要预测销售量为 30 万箱时，相对应的广告费用数值，需建立以销售量 y 为自变量，广告费用 x 为因变量的直线回归模型，然后将自变量数值代入方程，得到因变量的估计值，见表 8-5。

表 8-5　直线回归模型计算

编　号	广告费用(x)/万元	销售量(y)/万箱	y^2	xy
1	120	36.3	1 317.69	4 356
2	68.7	20.7	428.49	1 422.09
3	100.1	15.9	252.81	1 591.59
4	76.6	13.2	174.24	1 011.12
5	8.7	8.1	65.61	70.47
6	1	7.1	50.41	7.1
7	21.5	5.6	31.36	120.4
8	1.4	4.4	19.36	6.16
9	5.3	4.4	19.36	23.32
10	1.7	4.3	18.49	7.31
合计	405	120	2 377.82	8 615.56

根据最小平方法确定参数的公式，可计算出参数 c 与 d 为



$$\begin{cases} d=\dfrac{n\sum xy-\sum x\sum y}{n\sum y^2-(\sum y)^2} \\ c=\dfrac{\sum x}{n}-d\dfrac{\sum y}{n} \end{cases}$$

$$d=4.004\,6$$

$$c=-7.555\,2$$

将 c 和 d 代入回归方程，则 $x_c=-7.555\,2+4.004\,6y$，显然，此时的回归方程与原来的回归方程是两条不同的回归直线，具有不同的斜率和意义，只能给定自变量来推算因变量。当销售量为 30 万箱时，相对应的广告费用理论数值为

$$-7.555\,2+4.004\,6\times30=112.582\,8(\text{万元})$$

8.2.4　估计标准误差

根据直线回归方程，按给定的自变量值可以推算出相应的因变量值，即得出估计值 y_c。而估计值 y_c 与其对应的实际观察值 y 之间可能一致，也可能不一致，它们之间存在一系列离差，有的是正差，有的是负差。这些离差称为统计估计误差。这种估计误差的大小，可以说明推算结果的准确程度，即回归方程估计的准确程度。同时，统计上，一般是通过计算“估计标准误差”指标来反映回归方程的代表性的。若误差大，说明回归方程的代表性低；若误差小，则说明回归方程的代表性高。

估计标准误差(standard error of the estimate)是指因变量实际值与理论值离差的平均值，其计算原理与能够反映平均数代表性大小的标准差基本相同，定义公式为

$$S_{yx}=\sqrt{\frac{\sum(y-y_c)^2}{n-2}} \tag{8-10}$$



式中：S_{yx}为估计标准误差，其下标 yx 代表 y 依 x 而回归的方程；y_c为根据回归方程推算出来的因变量的估计值；y 为因变量的实际值；n 为数据的项数。

利用定义式计算估计标准误差，需要计算所有的估计值，计算量大，且计算比较麻烦。将定义式中 y_c用$a+bx$ 替代，经过化简，可得如下的计算公式：

$$S_{yx}=\sqrt{\frac{\sum y^2-a\sum y-b\sum xy}{n-2}} \tag{8-11}$$

【实例 8.5】 现仍以表 8－1 的有关资料为例，计算估计标准误差。

解：

根据表 8－1 和表 8－5 中的数据资料，可得如下计算结果：

$$\sum y^2=2\,377.82,\sum y=120,\sum xy=8\,615.56;$$

根据前面计算有 $a=4.07$，$b=0.1958$；由 $n=10$，有

$$S_{yx}=\sqrt{\frac{\sum y^2-a\sum y-b\sum xy}{n-2}}\approx 5.03$$

结果表明，国内 10 个品牌啤酒销售量的估计理论值与实际值的平均误差约为 5.03 万箱。由此可见，只有把回归估计值与估计标准误差结合起来分析运用，才更具有意义。

估计标准误差 S_{yx}与相关系数 r 在数量上也存在着密切的关系，可看成从另一个角度说明相关分析与回归分析之间的联系。两者之间的关系可由下列公式表述：

$$r=\sqrt{1-\frac{S_{yx}^2}{\sigma_y^2}} \tag{8-12}$$

$$S_{yx}=\sigma_y\sqrt{1-r^2} \tag{8-13}$$

从相互联系的两个算式中，可以看出 r 与 S_{yx} 的变化方向是相反的。当 r 越大时，S_{yx}越小，这时相关密切程度越高，回归直线的代表性越大；当 r 越小时，S_{yx} 越大，这时相关密切程度越低，回归直线的代表性越小；$r=\pm 1$ 时，$S_{yx}=0$，这时现象间完全相关，各相关点均落在回归直线上，此时对 x 的任何变化，y 总有一个相应的值与之对应；$r=0$ 时，S_{yx}取得最大值，这时现象间不存在直线关系。

8.2.5　判定系数

在回归分析中，除了可用估计标准误差反映回归方程估计的准确程度及回归直线代表性的大小外，还有一个广泛应用的指标，叫做判定系数(coefficient of determination)，它通常用 r^2 表示，用来测定回归方程拟合数据的好坏程度。判定系数 r^2 就是相关系数 r 的平方。当然，判定系数有它的基本公式：

$$r^2=\frac{\sum(y_c-\bar{y})^2}{\sum(y-\bar{y})^2} \tag{8-14}$$

式中：分母 $\sum(y-\bar{y})^2$ 是总离差平方和，分子 $\sum(y_c-\bar{y})^2$ 叫做回归平方和。

r^2越大，回归平方和在总离差平方和中所占的比重就越大，表明总离差中由回归方程解释的部分也越大，线性回归效果越好。

判定系数的范围在0～1。如果r^2为1，则表明两个变量之间有非常好的相关性，y的估计值与实际值之间没有差别，此时，估计标准误差为0；如果r^2为0，则回归方程不能用来预测y值，此时，估计标准误差最大。

例如，根据表8-1给出的2012年国内10个品牌啤酒广告费用及销售量，前面已计算出$r=0.885\ 6$，则判定系数$r^2=0.784\ 3$，这说明，总离差平方和中有78.43%可以用回归方程解释。

8.3　应用相关分析和回归分析应注意的问题

8.3.1　在定性分析的基础上进行定量分析

相关关系的有无、自变量和因变量的确定、相关的类型及计算结果的意义等问题，都必须根据人们对研究对象的了解，作出定性判断。因为相关分析的方法不能解释相关关系产生的原因，它本身不能判断现象之间是否存在质的关系，只有依靠定性分析并在此基础上进行定量分析，才能对现象之间的关系作出科学的判断。否则，若对本来没有内在关系的现象进行相关分析，将会导致虚假相关的错误。例如，研究某地区人口出生数量与该地区老年人口再婚数量之间的关系，显然是没有任何意义的。若据此继续进行回归分析并对现象进行推算预测，其结论将是荒谬的。

8.3.2　要注意现象质的界限及相关关系作用的范围

在进行相关分析与回归分析时，要注意现象质的界限及相关关系作用的范围。超出了这个范围，分析的结果就会与事实不符。特别是利用回归方程进行预测时，不能无限外推。因为，在预测中，指定x变量值未超出所用资料的数值范围，叫做内插回归预测，其预测效果好；预测中，指定x变量值超出所用资料的数值范围，叫做外推回归预测，预测效果随着x变量值对所用资料的数值范围的远离而逐渐降低。回归分析中，回归直线方程是采用最小平方法取得的，最小平方法指的是对现有资料范围配合一条“最佳”线，如果外推到范围以外，则不一定是最佳线了。根据样本数据所建立的回归方程，只有代表经济变量之间的数量条件才能够成立。如果忽视相关关系建立的条件，把这种关系无限制地向外推广，所得结论是不科学的。例如，商品的价格与商品的销售量只有在一定的范围内才具有负相关关系，商品的价格超过一定限度，销售量有可能与之作同方向变化。社会经济现象中，许多现象也如此。因此，用相关分析与回归分析方法进行推算和预测时，要注意它们的作用范围。

8.3.3　要将各种分析指标结合应用

相关分析与回归分析中的各种分析指标，既有区别又有联系。相关分析中的相关关系数，是反映现象之间直线关系密切程度的指标；回归分析中的直线回归方程，反映具有相关关系两变量之间数量的变化关系；而回归直线的代表性大小则由回归分析中的估计标准误差反映。它们分别从不同的层次和侧面揭示现象之间数量关系的特征。所以，在进行相



关分析和回归分析时，只有将它们有机地结合起来，才能更加全面、准确、科学地描述社会经济现象之间数量变化的规律性。

8.3.4 要尽可能使用大样本材料

大量观测是统计分析的基础，只有使用大样本资料进行大量观测，才能淡化偶然因素的影响，获得现象间的真实关系。不管是相关分析中通过计算相关系数判定相关密切程度，还是回归分析中利用回归方程进行统计预测，如果样本材料过少，都不易真实反映客观现象之间的关系。本章中给出的例题样本均较小，这只是为了写作上的方便，在实际中不宜采用。

本章小结

本章共有3节，安排3类内容。第一类是相关分析，包括相关关系的概念、种类，相关图表的绘制和相关系数的理论、计算；第二类是回归分析，包括回归分析的意义、特点，一元线性回归方程的建立、应用和估计标准误差的测定；第三类是应用相关分析和回归分析应注意的几个问题。

相关关系是现象之间确实存在的，但关系数值不确定的依存关系。相关关系从不同角度可以进行多种分类。其中，单相关是基础，线性相关是基本的相关。相关分析的主要对象是不完全的相关关系。

相关表是一种反映变量之间相关关系的统计表，相关图可以直观而形象地显示现象之间相关的性质和密切程度。相关表和相关图是进行相关分析的重要辅助工具。

相关系数是集中反映变量之间相关程度的代表性指标。相关系数的绝对值不超过1，它在-1与$+1$之间变动，相关系数绝对值越靠近1，表示x与y直线相关程度越高；越靠近0，表示x与y直线相关程度越低。相关系数的公式有定义式和计算式两种，实践中常采用计算式。配合直线回归方程$y_c=a+bx$的关键是找到合适的a和b，使实际的y值与对应的理论y_c值的离差平方和最小，即为最小平方法回归直线方程。

参数a代表直线的起点值，参数b称为回归系数，表示自变量x增加一个单位时，因变量y的平均增加值。

根据这个方程，只能在给定的“x”值时预测“y”的值，如果想了解当“y”已知，对应的“x”值为多少，则不能进行反推运算，而应求解出与原方程不一致的另一直线回归方程，回归方程的代表性可以通过估计标准误差指标来衡量。估计标准误差的公式有定义式和计算式两种，实践中常采用计算式。估计标准误差S_{yx}的数值越小，表示估计的准确性越高，回归方程的代表性越大。

在进行相关分析与回归分析时，应该在定性分析基础上进行定量分析；尽可能选择大样本资料进行分析；要注意现象质的界限及相关关系作用的范围；还应将相关系数、直线回归方程、估计标准误差等各种分析指标结合应用。只有这样，才能更加全面、准确、科学地描述社会经济现象之间数量变化的规律性。



应用与拓展

二手车车龄与售价关系的相关分析与回归分析

二手车的车龄是评估二手车价值的重要参考因素。大家都易相信，车龄与售价之间有着非常密切的关系，那就是，车龄越长，售价便越低。这里需要验证车龄与售价的相关性，来确定以上的结论。而且，还想知道，一般来说，对于不同品牌和系列的二手车，它们车龄每增加1年，汽车分别贬值多少？特别地，这项工作对评估师的正确估价有重要参考意义。

Lam先生是一家大型二手车连锁经营有限公司的评估师。他对大众(Volkswagen)品牌的高尔夫(Golf)系列车情有独钟，在工作中，他积累了大众高尔夫的二手车龄与最终成交的售价信息。他希望运用掌握的统计知识对搜集到的信息资料进行分析，得到科学的量化数字，使自己的工作更出彩。

大众高尔夫二手车的车龄和售价见表8-6。

表8-6　大众高尔夫二手车的车龄和售价

车龄/年	1	2	3	4	4	6	6	7	9	10
价格/千美元	27	27	21	26	20	23	20	12	10	14

根据以上资料，Lam先生绘制了反映二手车车龄与售价关系的散点图，如图8.7所示。又经过计算相关系数，得出车龄与售价之间线性相关系数为−0.856 9。可见，两者之间的确存在着高度负相关关系。

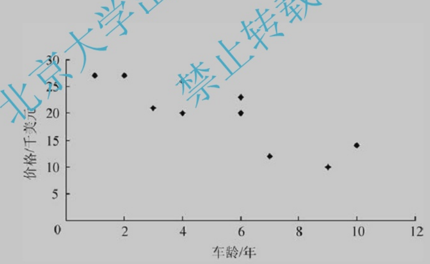


图8.7　大众高尔夫车龄与价格相关关系

在此基础上，Lam先生进行回归分析。以车龄为自变量x，价格为因变量y，则可以配合价格y与车龄x的直线回归方程为$y_c=29.381\,3-1.804\,1x$，其中，回归系数为−1.804 1，这表明，大众高尔夫车的车龄每增加1年，价格一般平均减少1 804.1美元。对于一辆常规使用的5年车龄的高尔夫，其价格的预测值是20 360.8美元。

有了这些结论作为依据，再结合自己丰富的汽车相关知识，Lam先生给很多客户提出了中肯的建议，帮助许多客户完成了高性价比的交易，工作受到了大家普遍好评。





习题与实训

一、单项选择题

1. 复相关是(　　)。

A. 两个或两个以上变量的相关关系

B. 3个或3个以上变量的相关关系

C. 正相关

D. 负相关

2. 如果相关系数 $r=-0.8901$，说明两个变量之间(　　)。

A. 完全相关　　B. 完全不相关　　C. 高度相关　　D. 显著相关

3. 变量之间的相关程度越低，则相关系数的数值(　　)。

A. 越接近于-1　　B. 越接近于$+1$　　C. 越接近于0　　D. 越小

4. 学院对学生考试成绩(百分制)与学习时间进行测定分析，建立了考试成绩与学习时间的直线回归方程：$y_c=160-2.3x$，该方程明显有错，错误在于(　　)。

A. 自变量和因变量的关系弄错了

B. a 值的计算是正确的，b 值是错误的

C. b 值的计算是正确的，a 值是错误的

D. a 值和 b 值的计算都有错误

5. 某地居民货币收入与购买商品支出相关资料见表8-7。

表8-7　某地居民货币收入与购买商品支出相关资料

年　份	2005	2006	2007	2008	2009	2010	2011	2012
货币收入/万元	36	37	38	40	42	44	47	50
购买商品支出/万元	30.0	31.0	32.0	33.2	34.8	36.5	39.0	41.6

则不需要计算，可判断(　　)可能是该地区2005—2012年购买商品支出依居民货币收入的回归方程。

A. $y_c=0.8532-0.8122x$　　B. $y_c=0.8532+0.8122x$

C. $y_c=40.8122-0.8532x$　　D. $y_c=40.8122+0.8532x$

6. 已知变量 x 和 y 的协方差为-67，x 的方差为192，y 的方差为31，其相关系数为(　　)。

A. 0.87　　B. -0.87　　C. 0.13　　D. -1.01

7. 若已知某企业相关资料见表8-8。

表8-8　某企业相关资料

产量/千件	2	3	4	3	4	5
单位产品成本/(元/件)	73	72	71	73	69	68

则不需要计算，便知道该企业产量与单位产品成本的相关系数为(　　)。

A. 0.908 7　　B. −0.909 1　　C. 1.032 3　　D. −0.499 8

8. 下列直线回归方程中，(　　)肯定是错误的。

A. $y=35+0.3x$，$r=0.8$　　B. $y=-124+1.4x$，$r=0.89$

C. $y=18-2.2x$，$r=0.74$　　D. $y=-87-0.9x$，$r=-0.9$

9. 下面变量之间的相关程度高的是(　　)。

A. 商品的广告支出费用与商品的销售收入的相关系数是 0.89

B. 汽车的维护费用与行驶里程的相关系数是 0.902 1

C. 警务支出费用与犯罪率的相关系数是−0.876 7

D. 平均流通费用率与商业利润率的相关系数是−0.94

10. 判定系数与相关系数、估计标准误差三者的关系是(　　)。

A. 估计标准误差越大，判定系数越大

B. 判定系数为 1 时，相关系数最小

C. 相关系数为−1 时，判定系数最大

D. 判定系数为 0 时，估计标准误差最小

二、多项选择题

1. 下列属于正相关的现象是(　　)。

A. 家庭收入越多，其消费支出越多

B. 生活水平越高，恩格尔系数越小

C. 航班正点率和顾客投诉次数之间的关系

D. 化妆品的广告费用和销售量的关系

E. 就餐账单的数额和服务费之间的关系

2. 关于相关系数，下面正确的是(　　)。

A. 相关系数的绝对值不超过 1

B. 相关系数是集中反映变量之间相关程度的代表性指标

C. 相关系数绝对值越靠近 1，表示 x 与 y 直线相关程度越高

D. 相关系数绝对值越靠近−1，表示 x 与 y 直线相关程度越低

E. 相关系数可正可负

3. 某乡相关资料见表 8－9。

表 8－9　某乡相关资料

耕地按耕作深度分组/cm	地块数/块	平均收获率/(kg/亩)	耕地按耕作深度分组/cm	地块数/块	平均收获率/(kg/亩)
10～12	7	200	16～18	12	310
12～14	10	230	18～20	5	340
14～16	16	270			

则(　　)。

A. 平均收获率与耕作深度之间存在着正相关关系



B. 平均收获率与耕作深度之间不一定存在着相关关系

C. 研究平均收获率与耕作深度之间的关系采用一元线性回归分析方法

D. 研究平均收获率与耕作深度之间的关系采用多元回归分析方法

E. 平均收获率与耕作深度之间的关系是复相关关系

4. 可以用来判断现象之间相关方向的指标有(　　)。

A. 估计标准误差　B. 判定系数　C. 相关系数

D. 回归系数　E. 两个变量的协方差

5. 在直线回归方程中(　　)。

A. 一个回归方程只能做一种推算

B. 两个变量中须确定自变量和因变量

C. 回归系数只能取正值

D. 回归系数与相关系数的符号是一致的

E. 同一方程中，给定 x 的值可以推测 y，给定 y 的值可以反过来推测 x

6. 进行相关分析和回归分析时，要注意的问题有(　　)。

A. 先进行定量分析然后进行定性分析

B. 要注意现象质的界限及相关关系作用的范围

C. 将相关系数、直线回归方程、估计标准误差等各种分析指标结合应用

D. 为计算方便，尽量选择小样本资料

E. 利用回归方程进行预测时，不能无限制地向外推测

7. 回归直线代表性较好的情况有(　　)。

A. 估计标准误差为 0　B. 相关系数的绝对值接近于 1

C. 判定系数为 0　D. 判定系数为 1

E. 相关系数为 0

8. 下列关于判定系数的描述正确的是(　　)。

A. 判定系数 r^2 就是相关系数 r 的平方

B. 判定系数的范围为 0～1

C. 如果 r^2 为 1，则表明两个变量之间有非常好的相关性

D. 如果 r^2 为 0，则估计标准误差为 0

E. 如果 r^2 为 0，估计标准误差最大

9. 在回归分析中，就两个相关变量 x 和 y 而言，由变量 y 依变量 x 回归和由变量 x 依变量 y 回归所得到的回归方程是不同的，这种不同表现在(　　)。

A. 方程中参数的数值不同　B. 方程中参数估计的方法不同

C. 估计标准误差的计算方法不同　D. 估计标准误差的数值不同

E. 参数表示的实际意义不同

10. 根据某地区历年人均收入 x(单位：元)与商品销售额 y(单位：万元)资料计算的有关数据如下：$n=9$；$\sum x=546$；$\sum y=260$；$\sum x^2=343\ 62$；$\sum xy=16\ 918$，则(　　)。

A. 以商品销售额为因变量的直线回归方程是 $y_c=26.92+0.92x$

B. 以商品销售额为因变量的直线回归方程是 $y_c=-26.92+0.92x$

C. 当人均收入每增加 1 元时，商品销售额平均增加 0.92 万元

D. 当人均收入为 1 000 元时，商品销售额为 920 万元

E. 当人均收入为 500 元时，商品销售额为 433.08 万元

三、判断题

1. 正相关指的是两个相关变量的数量变动方向都是上升的。（　　）

2. 相关系数越小，两变量之间相关的密切程度越低。（　　）

3. 只有当相关系数接近+1 时，才能说明两变量之间存在高度相关关系。（　　）

4. 若直线回归方程为 $y_c=120-23x$，则变量 x 与 y 之间存在负的相关关系。（　　）

5. 估计标准误差指标可以反映回归方程的代表性。若误差大，说明回归方程的代表性高；若误差小，则说明回归方程的代表性低。（　　）

6. (甲)某产品产量与单位成本的相关系数是−0.898 8；(乙)产品单位成本与利润率的相关系数是−0.956 7。因此，(乙)比(甲)的相关程度高。（　　）

7. 当相关系数 r 的绝对值等于 0 时，只能说明 x 与 y 两个变量之间无线性相关，不能说明两个变量不相关。（　　）

8. 判定系数越大，表明线性回归效果越好。（　　）

9. 在进行相关分析与回归分析时，尽量选择小样本资料，以利于计算。（　　）

10. 回归系数 b 与相关系数 r 不仅符号完全一致，而且存在数量关系。（　　）

四、填空题

1. 为了了解公司员工的工龄与其工作效率之间的相关性，公司人力资源部门进行抽样研究，8 位员工工龄(年)分别为 1、20、6、8、2、1、15、8，相应的效率分数分别是 6、5、3、5、2、2、4、3，则相关系数是______，相关程度是______。

2. 已知直线回归方程 $y_c=a+bx$ 中，$b=17.5$；又知 $n=30$，$\sum y=13\ 500$，$\bar{x}=12$，则可知 $a=$________。

3. 已知 x 变量的标准差为 2，因变量的标准差为 5，两变量的相关系数为 0.8，则回归系数是________。

4. 直线回归方程 $y_c=a+bx$ 中的参数 a 表示的意思是________，b 是________，其含义是________，估计待定参数使用的方法是________。

5. 相关分析是回归分析的________，而回归分析是相关分析的________。

6. 标准差是说明________的代表性的，而估计标准误差是反映________的代表性大小的。

7. 根据 49 个学生的中文成绩和英文成绩进行计算，中文成绩的标准差为 9.75 分，英文成绩的标准差为 7.9 分，两种成绩的协方差为 72 分，则中文成绩和英文成绩的相关方向是________，相关系数是________，相关程度是________。

8. 已知某一回归模型的判定系数为 0.64，则它的相关系数是________。

9. $n=6$，$\sum x=21$，$\sum y=426$，$\sum x^2=79$，$\sum y^2=30\ 268$，$\sum xy=1\ 481$，则变量 x 与变量 y 间的相关系数是________，回归系数是________。



10. 在线性相关条件下，自变量的标准差为10，因变量的标准差为16，相关系数为0.9，则回归系数为________。

五、应用能力训练题

1. 随机抽取10家航空公司，对其最近一年的航班正点率和顾客投诉次数进行了调查，所得数据见表8－10。

表8－10　被抽取的航空公司相关调查资料

航班正点率	81.8%	76.6%	76.6%	75.7%	73.8%	72.2%	71.2%	70.8%	91.4%	68.5%
顾客投诉次数/次	21	58	85	68	74	93	72	122	18	125

（1）编制相关表，绘制相关图，说明航班正点率和顾客投诉次数之间的关系。

（2）计算相关系数。

（3）用航班正点率作为自变量，顾客投诉次数作为因变量，求出估计的回归方程，并解释回归系数的意义。

（4）如果航班正点率为80%，估计顾客的投诉次数。

2. 航空公司经过统计调查得知，波音737飞机500英里飞行成本与乘客数量之间存在相关关系，调查数据资料见表8－11。

表8－11　航空公司所调查的相关数据资料

乘客数/人	61	63	67	69	70	74	76	81	86	91	95	97
成本/千美元	4.28	4.08	4.42	4.17	4.48	4.30	4.82	4.70	5.11	5.13	5.64	5.56

（1）建立飞行成本y与乘客数量x的直线回归方程。

（2）飞行每增加1名乘客，飞行成本将增加或减少多少美元?

（3）在此案例中，即使飞行时不搭载乘客，飞行500英里的飞行成本为多少美元?

3. 在某些国家和地区，在饭店就餐有给服务员小费的习惯。几乎所有人都相信，在账单的数额和小费之间存在着联系，即较大的账单就会有较多的小费。现搜集到6次午餐会的成对数据，见表8－12。

表8－12　6次午餐会的成对数据

账单/美元	33.46	50.68	87.92	98.84	63.60	107.34
小费/美元	5.50	5.00	8.08	17.00	12.00	16.00

（1）是否有足够的证据判断：在账单数额和小费数额之间存在某种联系?（它们相关的程度如何?）

（2）若存在某种联系，则使用这种联系来确定当账单数额为100美元时，应该留下多少小费?

（3）现已知给出小费10元，请推测其消费的账单数额为多少。

4. 科新公司科研经费支出与公司利润见表 8－13。

表 8－13　科新公司科研经费支出与公司利润

科研支出/万元	50	110	40	50	30	20
利润/万元	310	400	300	340	250	200

根据上述资料，进行以下计算。

(1) 计算科研经费支出与公司利润之间相关系数。

(2) 确定利润 y 与科研经费支出 x 的直线回归方程。

(3) 当科研经费支出为 120 万元时，利润估计为多少?

(4) 当利润为 380 万元时，科研经费支出大概是多少?

(5) 计算估计标准误差。

5. 投资分析家们通常认为债券的利率与最低贷款利率一般成反方向变化，即当贷款利率下降时，债券投资表现较好，而当贷款利率上升时，债券表现并不理想。能否根据最低贷款率的大小来预测债券利率? 请根据表 8－14 中给出的资料，判断两者之间的相关方向和相关密切程度。

表 8－14　债券利率与最低贷款利率

债券利率	5%	12%	9%	15%	7%
最低贷款利率	16%	6%	8%	4%	7%

若两者存在紧密关系，则确定出两者之间的直线回归方程，并预测，当最低贷款利率为 5%时，债券利率预测值为多少?

6. 根据 8 起火灾损失额 y(万元)及其火灾地点与消防队距离 x(km)的资料，经计算得知：$\sum x = 36.4, \sum y = 880, \sum x^2 = 207.54, \sum xy = 4\,544.6, \sum y^2 = 104\,214$，则当火灾发生地点与消防队之间的距离每增加 1km 时，损失额的变化是多少?

7. 某地区家计调查资料得到，每户平均月收入为 6 800 元，标准差为 800 元，平均消费支出为 5 200 元，方差为 40 000，支出对于收入的回归系数为 0.2，则

(1) 收入与支出的相关系数是多少?

(2) 写出支出对于收入的回归方程。

(3) 当收入为 7 300 元时，消费支出额为多少?

8. 已知 x、y 两变量的相关系数 $r = 0.8, \bar{x} = 20$，y 的平均数为 50，y 的标准差是 x 的标准差的 2 倍，请计算出 y 对 x 的回归直线方程。

9. 为研究员工学习掌握新技能的规律情况，抽查 5 位员工，学习一项新技能的时间(h)分别是 4、6、7、10、13，其相应的成绩(分)分别是 40、60、50、70、90，则

(1) 写出成绩依学习时间的直线回归方程。

(2) 求出该方程的估计标准误差。



10. 将学生分为8～12人的小组，

(1) 记录下每个小组成员的身高，再测量出他们的体重并记录下来，在身高和体重之间存在一种关系吗？如果存在，这种关系是什么？

(2) 测量每个小组成员身高和手臂伸开的宽度。测量臂宽时，测试对象应站直，将手臂平伸，如同飞机上的机翼。在身高与臂宽之间是否存在相关性？如果存在，求出由臂宽计算身高的回归方程。

(3) 记录下每个小组成员在1分钟内心跳的次数及心率。再记录下他(她)的身高。在心率和身高之间存在一种关系吗？如果存在，这种关系是什么？

11. 在第2章应用能力训练题中，搜集到了所在班级同学的身高、体重原始数据，经过第3章的学习，形成条理的统计资料。请根据这些材料，利用本章学过的知识，判断所在班级同学身高和体重之间相关程度如何。是否近似直线相关？是哪条直线？

附录 A　正态分布概率表

t	$F(t)$	t	$F(t)$	t	$F(t)$	t	$F(t)$
0.00	0.000 0	0.32	0.251 0	0.64	0.477 8	0.96	0.662 9
0.01	0.008 0	0.33	0.258 6	0.65	0.484 3	0.97	0.668 0
0.02	0.016 0	0.34	0.266 1	0.66	0.490 7	0.98	0.672 9
0.03	0.023 9	0.35	0.273 7	0.67	0.497 1	0.99	0.677 8
0.04	0.031 9	0.36	0.281 2	0.68	0.503 5	1.00	0.682 7
0.05	0.033 9	0.37	0.288 6	0.69	0.509 8	1.01	0.687 5
0.06	0.047 8	0.38	0.296 1	0.70	0.516 1	1.02	0.692 3
0.07	0.055 8	0.39	0.303 5	0.71	0.522 3	1.03	0.697 0
0.08	0.063 8	0.40	0.310 8	0.72	0.528 5	1.04	0.701 7
0.09	0.071 7	0.41	0.318 2	0.73	0.534 6	1.05	0.706 3
0.10	0.079 7	0.42	0.325 5	0.74	0.540 7	1.06	0.710 9
0.11	0.087 6	0.43	0.332 8	0.75	0.546 7	1.07	0.715 4
0.12	0.095 5	0.44	0.340 1	0.76	0.552 7	1.08	0.719 9
0.13	0.103 4	0.45	0.347 3	0.77	0.558 7	1.09	0.724 3
0.14	0.103 4	0.46	0.354 5	0.78	0.564 6	1.10	0.728 7
0.15	0.119 2	0.47	0.361 6	0.79	0.570 5	1.11	0.733 0
0.16	0.127 1	0.48	0.368 8	0.80	0.576 3	1.12	0.737 3
0.17	0.135 0	0.49	0.375 9	0.81	0.582 1	1.13	0.741 5
0.18	0.142 8	0.50	0.382 9	0.82	0.587 8	1.14	0.745 7
0.19	0.150 7	0.51	0.389 9	0.83	0.593 5	1.15	0.749 9
0.20	0.158 5	0.52	0.396 9	0.84	0.599 1	1.16	0.754 0
0.21	0.166 3	0.53	0.403 9	0.85	0.604 7	1.17	0.758 0
0.22	0.174 1	0.54	0.410 8	0.86	0.610 2	1.18	0.762 0
0.23	0.181 9	0.55	0.417 7	0.87	0.615 7	1.19	0.766 0
0.24	0.189 7	0.56	0.424 5	0.88	0.621 1	1.20	0.769 9
0.25	0.197 4	0.57	0.431 3	0.89	0.626 5	1.21	0.773 7
0.26	0.205 1	0.58	0.438 1	0.90	0.631 9	1.22	0.777 5
0.27	0.212 8	0.59	0.444 8	0.91	0.637 2	1.23	0.781 3



续表

t	F(t)	t	F(t)	t	F(t)	t	F(t)
0.28	0.220 5	0.60	0.451 5	0.92	0.642 4	1.24	0.785 0
0.29	0.228 2	0.61	0.458 1	0.93	0.647 6	1.25	0.788 7
0.30	0.235 8	0.62	0.464 7	0.94	0.652 8	1.26	0.792 3
0.31	0.243 4	0.63	0.473 1	0.95	0.657 9	1.27	0.795 9
1.28	0.799 5	1.61	0.892 6	1.94	0.947 6	2.54	0.795 9
1.29	0.803 0	1.62	0.894 8	1.95	0.948 8	2.56	0.988 9
1.30	0.806 4	1.63	0.896 4	1.96	0.950 0	2.58	0.989 5
1.31	0.809 8	1.64	0.899 0	1.97	0.951 2	2.60	0.990 1
1.32	0.813 2	1.65	0.901 1	1.98	0.952 3	2.62	0.990 7
1.33	0.816 5	1.66	0.903 1	1.99	0.953 4	2.64	0.991 2
1.34	0.819 8	1.67	0.905 1	2.00	0.954 5	2.66	0.991 7
1.35	0.823 0	1.68	0.907 0	2.02	0.956 6	2.68	0.992 2
1.36	0.826 2	1.69	0.909 0	2.04	0.958 7	2.70	0.992 6
1.37	0.829 3	1.70	0.910 9	2.06	0.960 6	2.72	0.993 1
1.38	0.832 4	1.71	0.912 7	2.08	0.962 5	2.74	0.993 5
1.39	0.835 5	1.72	0.914 6	2.10	0.964 3	2.76	0.993 9
1.40	0.838 5	1.73	0.916 4	2.12	0.966 0	2.78	0.994 2
1.41	0.841 5	1.74	0.918 1	2.14	0.967 6	2.80	0.994 6
1.42	0.844 4	1.75	0.919 9	2.16	0.969 2	2.82	0.994 9
1.43	0.847 3	1.76	0.921 6	2.18	0.970 7	2.84	0.995 2
1.44	0.850 1	1.77	0.923 3	2.20	0.972 2	2.86	0.995 5
1.45	0.852 9	1.78	0.924 9	2.22	0.973 6	2.88	0.995 8
1.46	0.855 7	1.79	0.926 5	2.24	0.974 9	2.90	0.996 0
1.47	0.858 4	1.80	0.928 1	2.26	0.976 2	2.92	0.996 2
1.48	0.861 1	1.81	0.929 7	2.28	0.977 4	2.94	0.996 5
1.49	0.863 8	1.82	0.931 2	2.30	0.978 6	2.96	0.996 7
1.50	0.866 4	1.83	0.932 8	2.32	0.979 7	2.98	0.996 9
1.51	0.869 0	1.84	0.934 2	2.34	0.980 7	3.00	0.997 3
1.52	0.871 5	1.85	0.935 7	2.36	0.981 7	3.20	0.997 9
1.53	0.874 0	1.86	0.937 1	2.38	0.982 7	3.40	0.998 6
1.54	0.876 4	1.87	0.938 5	2.40	0.983 6	3.80	0.999 68

续表

t	F(t)	t	F(t)	t	F(t)	t	F(t)
1.55	0.878 9	1.88	0.939 9	2.42	0.984 5	4.00	0.999 94
1.56	0.881 2	1.89	0.941 2	2.44	0.985 3	4.50	0.999 993
1.57	0.883 6	1.90	0.942 6	2.46	0.986 1	5.00	0.999 999
1.58	0.885 9	1.91	0.943 9	2.48	0.986 9		
1.59	0.888 2	1.92	0.945 1	2.50	0.987 6		
1.60	0.890 4	1.93	0.946 4	2.52	0.988 3		

附录B　t-分布临界值表

$n-1$ \ $t_\alpha(n-1)$ \ α	0.25	0.10	0.05	0.025	0.01	0.005
1	1.000 0	3.077 7	6.313 8	12.706 2	31.820 7	63.657 4
2	0.816 5	1.885 6	2.920 0	4.320 7	6.964 6	9.924 8
3	0.764 9	1.637 7	2.353 4	3.182 4	4.540 7	5.840 9
4	0.740 7	1.533 2	2.131 8	2.776 4	3.746 9	4.604 1
5	0.726 7	1.475 9	2.015 0	2.570 6	3.364 9	4.032 2
6	0.717 6	1.439 8	1.943 2	2.446 9	3.142 7	3.707 4
7	0.711 1	1.414 9	1.894 6	2.364 6	2.998 0	3.499 5
8	0.706 4	1.396 8	1.859 5	2.306 0	2.896 5	3.355 4
9	0.702 7	1.383 0	1.833 1	2.262 2	2.821 4	3.249 8
10	0.699 8	1.372 2	1.812 5	2.228 1	2.763 8	3.169 3
11	0.697 4	1.363 4	1.795 9	2.201 0	2.718 1	3.105 8
12	0.695 5	1.356 2	1.782 3	2.178 8	2.681 0	3.054 5
13	0.693 8	1.350 2	1.770 9	2.160 4	2.650 3	3.012 3
14	0.692 4	1.345 0	1.761 3	2.144 8	2.624 5	2.976 8
15	0.691 2	1.340 6	1.753 1	2.131 5	2.602 5	2.946 7
16	0.690 1	1.336 8	1.745 9	2.119 9	2.583 5	2.902 8
17	0.689 2	1.333 4	1.739 6	2.109 8	2.566 9	2.898 2
18	0.688 4	1.330 4	1.734 1	2.100 9	2.552 4	2.878 4
19	0.687 6	1.327 7	1.729 1	2.093 0	2.539 5	2.860 9
20	0.687 0	1.325 3	1.724 7	2.086 0	2.528 0	2.845 3
21	0.684 4	1.323 2	1.720 7	2.079 6	2.517 7	2.831 4
22	0.685 8	1.321 2	1.717 1	2.073 9	2.508 3	2.818 8
23	0.685 3	1.319 5	1.713 9	2.068 7	2.499 9	2.807 3
24	0.684 8	1.317 8	1.710 9	2.063 9	2.492 2	2.796 9
25	0.684 4	1.316 3	1.708 1	2.059 5	2.485 1	2.787 4
26	0.684 0	1.315 0	1.705 6	2.055 5	2.478 6	2.778 7
27	0.683 7	1.313 7	1.703 3	2.051 8	2.472 7	2.770 7
28	0.683 4	1.312 5	1.701 1	2.048 4	2.467 1	2.763 3
29	0.683 0	1.311 4	1.699 1	2.045 2	2.462 0	2.756 4
30	0.682 8	1.310 4	1.697 3	2.042 3	2.457 3	2.750 0

附录 C　随机数表(摘录)

	112 345 678 910	212 345 678 910	312 345 678 910	412 345 678 910	512 345 678 910
1	6 119 690 446	2 645 747 774	5 192 433 729	6 539 459 593	4 258 260 527
2	1 547 445 266	9 527 079 953	5 936 783 848	8 239 610 118	3 321 159 466
3	9 455 728 573	6 789 754 387	5 462 244 431	9 119 042 592	9 292 745 973
4	4 248 116 213	9 734 408 721	1 686 848 767	0 307 112 059	2 570 146 670
5	2 352 378 317	7 320 889 837	6 893 591 416	2 625 229 663	0 552 282 562
6	0 449 352 494	7 524 633 824	4 586 251 025	6 196 279 335	6 533 712 472
7	0 054 997 654	6 405 188 159	9 611 963 896	5 469 282 391	2 328 729 529
8	3 596 315 307	2 689 809 354	3 335 135 462	7 797 450 024	9 010 339 333
9	5 980 808 391	4 542 726 842	8 360 949 700	1 302 124 892	7 856 520 106
10	4 605 885 236	0 139 092 286	7 728 144 077	9 391 083 647	7 061 742 941
11	3 217 900 597	8 737 925 241	0 556 707 007	8 674 317 157	8 539 411 838
12	6 923 461 406	2 011 745 204	1 595 660 000	1 874 392 423	9 711 896 388
13	1 956 541 430	0 175 875 379	4 041 921 585	6 667 436 806	8 496 285 207
14	4 515 514 938	1 947 607 246	4 366 794 543	5 904 790 033	2 082 669 541
15	9 486 431 994	3 616 810 851	3 488 881 553	0 154 035 456	0 501 451 176
16	9 808 624 826	4 524 028 404	4 499 908 896	3 909 473 407	3 544 131 880
17	3 318 516 232	4 194 150 949	8 943 548 581	8 869 541 904	3 754 873 043
18	8 095 100 406	9 638 270 774	2 015 123 387	2 501 625 298	9 462 461 171
19	7 975 249 140	7 196 128 296	6 986 102 591	7 485 220 539	0 038 759 579
20	1 863 332 537	9 814 506 571	3 101 024 674	0 545 561 427	7 793 891 936

注：此表是美国兰德公司 1955 年编制出版的随机数表中的 1 000 个数字。